

交互作用統計量に基づく消費者選好構造の研究

著者	石田 実
内容記述	筑波大学博士（経営学）学位論文・平成24年1月31日授与（乙第2572号）
発行年	2012
URL	http://hdl.handle.net/2241/121341

交互作用統計量に基づく消費者選好構造の研究

筑波大学審査学位論文（博士）

2 0 1 1

石 田 実

筑波大学大学院

ビジネス科学研究科 企業科学専攻

目 次

第 1 章	序論	4
1.1	研究の背景	4
1.2	研究の目的	6
1.3	研究の構成	9
第 2 章	先行研究のレビューと研究課題	11
2.1	研究のレビューの範囲	11
2.2	先行研究のレビュー	12
2.2.1	ピースミール処理とカテゴリーベース処理	12
2.2.2	カテゴリー概念とカテゴリーの類似関係	13
2.2.3	競争市場構造分析	15
2.2.4	消費者による製品選択の異質性の所在	16
2.2.5	製品カテゴリーの類似関係の 2 つの推定経路：知覚と行動	18
2.2.6	製品カテゴリーの類似関係を用いた製品のクラスター分析	20
2.2.7	消費者の選好構造の空間的表現	22
2.3	本研究における主要課題	23
第 3 章	行動面からの製品カテゴリーの類似関係の推定	29
3.1	はじめに	29
3.2	検証の手法	31
3.2.1	類似係数	31
3.2.2	クラスターバリデーション	35
3.2.3	クラスタリング手法	40
3.3	実証分析と結果	41
3.3.1	データ	41

3.3.2	再現性の検証	43
3.3.3	均等性の検証	43
3.3.4	外的基準の検証	47
3.3.5	内的基準の検証	48
3.3.6	解釈可能性の検証	49
3.3.7	結果	51
3.4	考察と課題	51
3.4.1	考察	51
3.4.2	課題	52
第4章	交互作用統計量に基づく製品選択の研究	54
4.1	はじめに	54
4.2	先行研究	55
4.3	モデル	58
4.3.1	モデルの意図	58
4.3.2	理論モデル	59
4.3.3	周辺確率の推定法 (1) - 定数と仮定	62
4.3.4	周辺確率の推定法 (2) - 消費者の購買数と製品の販売数が独立と仮定	63
4.3.5	結果の規格化	64
4.4	音楽CD販売POSデータを用いたモデルの検証	66
4.4.1	目的と手順	66
4.4.2	データ	67
4.4.3	結果	68
4.5	類似関係を利用するプロモーションへの応用	70
4.5.1	売上の交差弾力性と評価手順	70
4.5.2	評価結果	72
4.6	結論と今後の課題	74
4.6.1	結論	74
4.6.2	今後の課題	74

第5章	消費者の選好構造の空間的表現の研究	78
5.1	はじめに	78
5.2	先行研究と課題	80
5.2.1	先行研究	80
5.2.2	研究の課題	84
5.3	分析手法	84
5.3.1	消費者の選好構造の空間的表現の構成	84
5.3.2	消費者セグメントと製品カテゴリーの対応について	86
5.3.3	消費者の配置と選好の強さの検証法	86
5.3.4	新製品の新規採択者の配置の推移と売上推移パターン	88
5.4	実証分析	89
5.4.1	データ	89
5.4.2	分析結果1：消費者の配置と製品への選好の強さの関係評価	90
5.4.3	分析結果2：消費者の配置の分布と新製品の売上パターンとの関連を 評価	92
5.5	結果と応用および今後の課題	97
5.5.1	結果	97
5.5.2	プロモーション戦略への応用	98
5.5.3	学術的貢献と今後の課題	99
第6章	結論	101
6.1	結論	101
6.2	学術的貢献	103
6.3	実務的示唆	105
6.4	今後の課題	107
参考文献		
謝辞		
付録A		118
A.1	第3章の分析結果一覧	118

第1章 序論

1.1 研究の背景

社会が豊かになるにつれて、毎日の生活の中で時間の使い方の自由度が増し、TVや新聞などのマスメディアからの情報の他に、インターネットで主体的に情報探索する機会が増えて、個人の価値観やライフスタイルが多様化する傾向にある。価値観が多様化する消費者に対して製品やサービスの価値を伝えるためには、それぞれの消費者に適したプロモーションや製品を提供する仕組みが必要となり、従来の調査紙形式やインタビューによる購入意向調査に加えて、購買データを活用した調査分析の仕組みが模索されている。例えば、多くの小売店がポイントなどの特典のある顧客カードやクレジットカードを導入したり、電子マネーのように利便性の高い決済システムを導入して、顧客の囲い込みを行っているが、そのためのシステムインフラは、同時に個人のID付き購買履歴を収集する仕組みでもあり、顧客の囲い込み戦略のための分析インフラと一体化している。またインターネット上のオンラインショップでは、購買履歴やウェブサイトの閲覧履歴がシステムのログとして自動的に残るので、これらの消費者のIDの付いた膨大な履歴をデータベース化してマーケティングに活用している。そして、これらのデータベースを活用し、オンラインショップ上の製品推奨や検索連動型広告、電子メールやモバイル広告など、マスメディアとは異なる媒体を通したプロモーションが提案されている。

これらの行動履歴データの活用を、データの収集、運用と分析という段階で見ると、2010年時点でPOSシステムの累積導入店舗数は48万店（（財）流通システム開発センター調べ）となり、既に多くの小売店が個人履歴データを蓄積している段階にあると言える。次に、運用に向けた分析であるが、例えばポイントシステムは一種の割引であるため、コストに見合った囲い込みの効果を定量的にモニタリングする運用が求められる。そのための分析機能は、従来の小売店の業務には無かった機能なので、社内に分析体制を敷いて人件費負担を増やしたり、社外に重要な経営情報を渡して分析を外注するといった対応の、コストと効

果のトレードオフを多くの小売店は見極めようとしている。ID付きPOSデータを活用して個人の異質性を捉えるニーズに対しては、ベイズ推定を用いた手法が多数提案されているが（阿部・近藤, 2005; Rossi, Allenby & McCulloch 2006, 2006; 照井, 2008; 佐藤, 2010; 樋口, 2011）、これらの高度な分析を活用している企業は、経営体力のある大企業に限られているのではなかろうか。

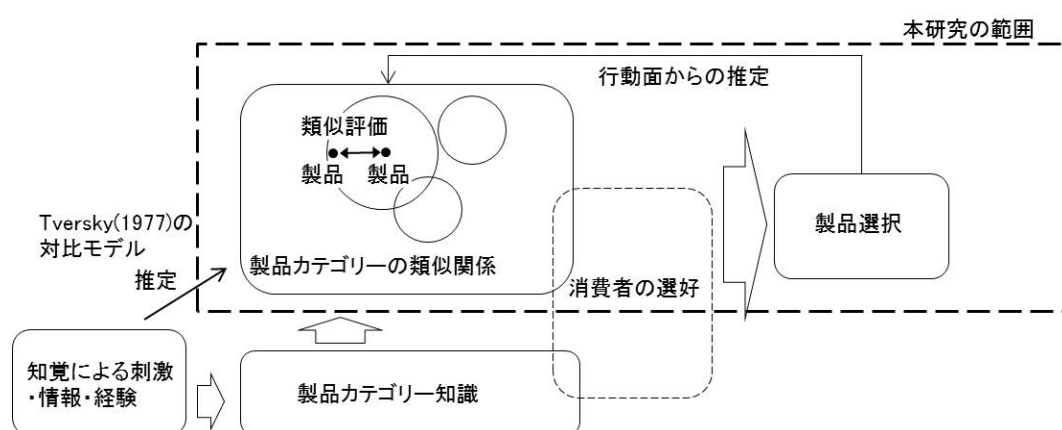
一方で、特定の市場における製品の特徴や消費者行動に関する知見を必要とせず、消費者の購買意思決定プロセスを明示的に扱わない購買履歴データの活用法として、協調フィルタリングを用いたレコメンドシステムがある。「この商品を買った人はこの商品も買っています。」といった製品推奨を行う協調フィルタリングは、Goldberg, Nichols, Oki & Terry (1992) がコンピュータサイエンスの分野で提案したアルゴリズムで、基本的な構成として推定を必要としないため、運用が容易な手法である。協調フィルタリングの応用としてのレコメンドシステムの効果については、消費者心理の観点から、消費者の情報探索コストを軽減し、多数の購入候補の中から多くの情報を処理してかしこい購入決定をしたという満足感を与えるという指摘がある（Haubl & Trifts, 2000; Swaminathan, 2003; Haubl & Murray, 2003; Lurie, 2004; Aksoy, Bloom, Lurie & Cooil, 2006）。協調フィルタリングの手法が普及した背景には、レコメンドシステムの有用性と、マーケティングの観点からの分析を必要としない導入コストの低さがあったと考える。

協調フィルタリングの手法の限界としては、市場の競争構造を捉えるのに適さない事がある。マーケティングのプロモーションに携わる方から、「行動データの活用に対するニーズは強い。しかし、製品カテゴリーや消費者セグメントを無視するシステムの手法は現場になじまない。」というコメントを頂戴する事も多い。システムとしての協調フィルタリングは、消費者の購買プロセスを捉える意図で考案されたものではないので、サービスとしてのレコメンドシステムへの実装以外でマーケティングと関連が薄いのは当然である。しかし、類似評価を利用した協調フィルタリングのアイデアが提出される1992年（Goldberg, et al., 1992）より15年早く、Tversky（1977）は認知心理学の研究領域において、類似評価を用いたカテゴリーベース処理のモデル（対比モデル: Contrast Model）を提案している。カテゴリーベース処理では、人が様々な情報や製品をカテゴリー化することでその情報や製品を理解し、情報処理をおこなうという考え方に立っている（新倉, 2005）。レコメンドシステムの研究ではTversky（1977）の引用を見ないため、これらの研究分野では消費者の情報処理プロセスに対する関心は低いと思われるが、ともに類似概念による定量的な評価を用いてい

る点が興味深い。

ライフスタイルが多様化すれば、個々の消費者が自身の基準で様々な情報を判断することが多くなり、複雑な情報処理を単純化し、認知努力を節約する仕組みとして知られるカテゴリーベース処理に頼る状況が多くなろう。清水（1999）は、今日では消費者の関与が高く情報収集の動機が高い場合を除いて、カテゴリーベース処理のための情報収集・処理が行われていると指摘している。個々の消費者は、ライフステージなどの属性が同じであっても、異なった経験を有し、それぞれが異なるカテゴリー知識を持っていると考えられる。あるいは、カテゴリー知識を共有しながら、異なるカテゴリーへの選好を強めているとも解釈できよう。行動面のデータから消費者セグメントや製品カテゴリーを捉える仕組みは、十分に研究されているとはいいがたく、類似概念に着目した研究は、ID付きPOSデータの蓄積を行っている小売店のニーズへの対応に役立つ可能性がある。行動面のデータをTverskyの対比モデルを踏まえて解釈することは、新たな知見を見出す手段になると期待する。多様化する消費者の選好を捉える手法として、ベイズモデリングなどの高度な手法に加え、Tverskyの対比モデルが提示する類似評価の活用法への関心が、本研究の動機となっている。

1.2 研究の目的



消費者の選好と、製品カテゴリーの類似関係および製品カテゴリー知識との包含関係・独立性については、議論の余地があるため本研究を通して考察する。

図 1-1. 製品カテゴリーの類似関係の位置づけ

前節の研究動機を踏まえ、本研究では、消費者が内面に持つ製品カテゴリーの類似関係を、行動面から推定する研究を行う。ここで製品カテゴリーの類似関係とは、Tversky の対

比モデルと同様に、製品間の連続的な近接関係であり、研究対象とする市場の任意の製品 a と b の組み合わせに対して、消費者にとっての類似の強さを表す関数

$$s : (a, b) \mapsto \text{実数値} \quad (1.1)$$

と定める。関数 $s(a, b)$ が満たすべき条件を対比モデルとして提示した Tversky(1977) は、図 1-1 の概念図に示すように、人が対象や情報を知覚することで、それらをカテゴリー化して理解するためのカテゴリー知識を得るとして、人の知覚情報からカテゴリー構造を評価する関数 $s(a, b)$ の一般式を示した。本研究では、消費者の内面にあるカテゴリー構造が製品選択に影響しているとの見通しの下に、消費者の製品選択の行動結果からカテゴリー構造を表現する類似評価の関数を推定する。また、製品カテゴリーの類似関係に対応した消費者セグメントを推定し、消費者の選好構造を捉える事を目指す。なお、消費者の製品の知覚情報や、それらの情報や購買経験によって形成される製品カテゴリー知識については、本研究の範囲外とする。

研究目的として、競争市場構造分析の手法を踏まえて、市場の製品カテゴリーと消費者セグメントの対応として市場構造を捉え、これにより新製品の売上パターンに関する知見を導いて、マーケッターがプロモーション戦略を評価する手法を提案する。まず、消費者の製品カテゴリーの類似関係を定量的に推定する事を目指す。次に、推定した製品カテゴリーの類似関係から製品選択に至る関係を捉え、製品カテゴリーの類似関係と製品選択との関係を明らかにする事を目指す。製品カテゴリーの類似関係の定量的評価については、Tversky (1977) の対比モデルとそれに続く多数の研究があるが、これらのモデルは消費者の知覚に関する質問項目からの評価であるため、行動面からの評価法を示して、その妥当性を明らかにする。また、行動面から推定した製品カテゴリーの類似関係と、これに対応する消費者セグメントを空間的に表現し、競争市場構造分析の空間的表現として多様な消費者の選好構造を捉える。特に新製品の売上パターンを左右する要因を見出し、プロモーション戦略の判断材料として示す。以上の研究目的の流れを図 1-2 に示す。

第 1 の研究目的として、行動面から製品のカテゴリーの類似関係を定量的に推定する手法を提案し、その妥当性を検証する。カテゴリーの類似関係の推定手法としては、知覚データを扱う Tversky (1977) の対比モデルを行動データに適用する方針のもとに、対比モデルで示された一般的な類似評価の算式から具体的な算式を選択する。行動面からカテゴリーの類似関係が推定できると考える根拠は、第 1 に、カテゴリー知識形成の直接的な刺激である知

覚データを用いて、カテゴリーの類似関係を評価する妥当性を Tversky (1977) が示した事にある。第2に、結果としての行動から要因としてのカテゴリーの類似関係を推定できると考えるためである。推定したカテゴリーの類似関係は製品×製品の類似評価を値とする行列表現となるが、行列表現のままでは推定の妥当性が評価できない。そこで、類似評価値に対してクラスター分析を適用し、クラスター分析の妥当性評価（クラスターバリデーション）の手順に従って、行動面から消費者のカテゴリーの類似関係を推定する妥当性を検証する。

第2の研究目的として、行動面から推定した製品カテゴリーの類似関係が、消費者の製品選択に与える影響を明らかにする。行動面のデータから類似評価を用いて製品選択を説明する手法としては、オンラインショップのレコメンドシステムに利用される協調フィルタリングが知られている。その有用性は、今日のレコメンドシステムの普及により広く明らかと言えるが、コンピュータサイエンスの領域のアルゴリズムであるために、製品カテゴリーや消費者セグメントの概念を介さない計算手続きとなっている。また、消費者の異質性の設定も明らかでなく、マーケティングに活用する手法としては不十分と言える。そこで、カテゴリーの類似関係から製品選択を説明するモデルを新たに構成し、モデルを用いて消費者の製品選択における製品カテゴリーの類似関係の役割や、消費者の異質性の所在を議論する。

第3の研究目的として、まず、競争市場構造分析の手法に従い、製品カテゴリーの類似関係に対応した消費者セグメントの推定を行って、その推定法の妥当性を評価する。具体的には、製品のカテゴリーの類似関係を推定するのと同様の手法により、消費者セグメントの類似関係を評価できることを示す。検証手続きは、製品カテゴリーと消費者セグメントとの対応を、離散的な分類ラベルによる表現と、空間的な表現の2通りで確認する。分類ラベルによる検証は、クラスターバリデーションを行う第1の研究目的の際に併せて行う。空間的表現については、製品と消費者を同じ空間に配置し、消費者が製品から遠ざかるにつれて購買確率が低下することを確認する。空間的表現には、消費者と製品の関係を連続に捉えられる利点があり、行動面のデータのタイムスタンプを用いれば、市場における動的な変化を捉えることができる。空間的に表現されたデータは、そうでないデータよりも適用可能な統計手法の範囲が広いので、分析のための表現形式として利便性が高い。実証として、市場に投入された新製品を購入する消費者の空間的な推移を評価し、新製品の売上パターンを左右する要因についての新たな知見を見出して、プロモーション戦略への応用を提案する。

これらの3つの研究により、製品カテゴリーの類似関係が、消費者の製品選択において果たす役割を明らかにし、類似概念を用いた市場の理解や、プロモーション戦略への応用を

示す。

1.3 研究の構成

本研究の構成は以下の通りとする。

第2章では、本研究に関連する最新の研究動向を整理し、それらの研究分野における本研究の位置付けを明らかにする。また、関連する研究分野における問題点を検討し、本研究で取り組むべき課題を明らかにする。

第3章では、Tverskyの対比モデルを踏まえて、行動面から製品カテゴリーの類似関係を定量的に推定する手法を提示して、その妥当性を検証する。検証手法としては、候補となる類似評価式毎にクラスター分析を適用し、クラスターバリデーションの基準で相対的な優劣を評価する。対比モデルでは、統計分野において類似係数と呼称される指標群の一般式として類似評価式を与えている事を踏まえて、主要な類似係数を類似評価式の候補とする。類似係数の優劣の判定では、クラスター分析の分類結果は教師なし分類となり絶対的な評価基準が存在しないので、多面的に相対的な評価を注意深く行う。ただし、単なる網羅的な評価とならないように、類似係数の優劣の評価軸として交互作用と主効果を設定し、望ましい類似評価の特徴について明らかにする。

第4章では、第3章で明らかにする製品カテゴリーの類似関係の推定結果を踏まえて、製品カテゴリーの類似関係と製品選択との関係を明らかにする。具体的には、類似係数を用いて消費者の製品選択を説明するモデルを構成し、その妥当性を検証する。評価基準として、実データと模擬データを用い、推定に用いた行動データより将来の購買を有意に予測できることを示す。提案モデルの構築を通して、製品選択における消費者の異質性の所在を明らかにし、製品カテゴリーの類似関係が売上に与える影響を評価する。

第5章では、競争市場構造分析の手法に従い、行動面のデータを用いて製品カテゴリーに対応した消費者セグメントを推定する手法を示し、その空間的表現の妥当性を評価する。具体的には、製品と消費者を同じ空間に布置し、消費者が製品から遠ざかるにつれて購買確率が低下することを確認して、消費者セグメントが製品カテゴリーに対応している事と、空間的表現の妥当性を確認する。特に新製品について、新規採択者の布置の推移を評価することで、競争市場構造分析の観点から新製品のプロモーション戦略に関する知見を見出し、本分析手法の有用性を確認する。

第6章では、本研究における3つの研究の成果を整理し、学術的貢献と実務的示唆について述べる。また、今後の残された課題について考察する。

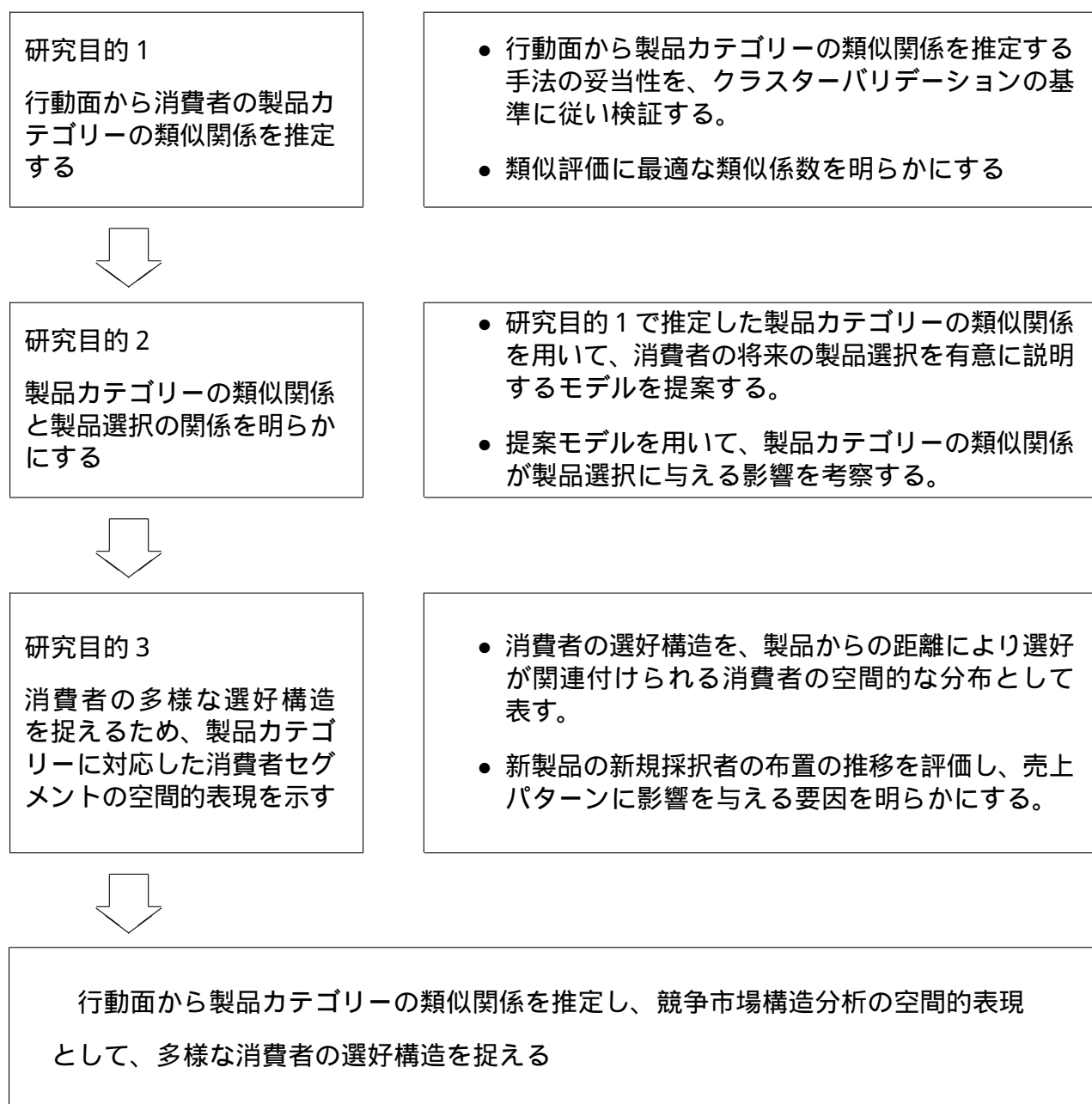


図 1-2. 3つの研究目的

第2章 先行研究のレビューと研究課題

2.1 研究のレビューの範囲

本章では、カテゴリーベース処理に着目した製品選択を研究する際に参考となる先行研究のレビューを行う。まず、3つの研究課題に取り組むために必要となる先行研究のレビューの範囲を設定する。

研究課題1では、消費者の行動面から製品カテゴリーの類似関係を推定する手法を提案し、その妥当性をクラスターバリデーションの手続きに沿って検証する。この手法は、製品カテゴリーの類似関係が購買行動に与える影響に着目した手法である事を踏まえて、カテゴリー概念と、消費者のカテゴリー知識を用いた製品理解であるカテゴリーベース処理と、製品属性の便益評価に基づく情報処理であるピースミール処理に関する研究についてレビューする。また、製品カテゴリーの類似関係を消費者の製品に対する知覚を用いて説明した Tversky(1977) の対比モデルと、関連する研究についてレビューする。類似関係の評価式としての類似係数の選択は、本研究の分析手法の根幹をなす研究テーマである事を踏まえ、類似係数と検証手法のクラスターバリデーションについてレビューする。研究課題2では、製品カテゴリーの類似関係から製品選択を説明する関係式を導き、行動面から推定した製品カテゴリーの類似関係が、将来の消費者の製品選択を有意に説明する事を検証するのを踏まえて、類似関係で消費者の購買を説明する強調フィルタリングのアルゴリズムとレコメンドシステムに関する研究についてレビューする。また、その関係式の構築を通して、消費者の製品選択に異質性を与える要因を明らかにするので、マーケティングにおける異質性の所在の扱いと、選好構造の表現形式についてレビューする。研究課題3では、製品カテゴリーの類似関係に対応した消費者セグメントを推定し、製品の競合と消費者の選好の対応を空間的に捉える競争市場構造分析のフレームを提示する事を踏まえて、競争市場構造分析の研究についてレビューする。

2.2 先行研究のレビュー

2.2.1 ピースミール処理とカテゴリーベース処理

消費者が購買決定に際して行う情報処理の代表的なものに、表 2.1 の通り、ピースミール処理とカテゴリーベース処理がある (Sujan, 1985; Meyers-levy & Tybout, 1989)。

表 2-1 . ピースミール処理とカテゴリーベース処理

	ピースミール処理	カテゴリーベース処理
概要	製品を属性に分解して処理	製品をカテゴリーに照らして理解
構造	断片的・精緻	包括的
処理の負荷	高負荷、情報収集コスト高い	低負荷、情報収集コスト低い
関与との関係	高関与時	低関与時

前者のピースミール処理は、製品を便益と関連付けられる属性の束と捉え、製品を構成する属性に分解して処理する手続きである。この情報処理を捉える分析手法として多属性分析 (中西, 1984) があり、また、具体的な定量モデルとして、多属性型態度モデル (Fishbein, 1963) がある。ピースミール処理では、製品を多数の属性に分解し、属性毎に便益を評価した後に再統合して総合的な製品の評価を求めるので、合理的で複雑な情報処理が必要となる。また、製品の属性評価のための情報を入手する情報探索コストも大きい。このため、清水 (1999) は、これまで多属性型態度モデルが盛んに議論されてきたが、消費者の関与が高く情報収集の動機が高い場合を除いて、今日ではカテゴリーベース処理のための情報収集・処理が行われていると考えられていると述べている。

カテゴリーベース処理は、提示された製品や情報を事前に消費者が知識としてもつ何らかの基準で作られたカテゴリーと照らし合わせることによって、対象を理解しようとする方式である。製品属性が製品選択の手がかりにならない製品の評価では、ピースミール処理による評価が困難となるためにカテゴリーベース処理の有用性は高まると考えられる。新倉 (2005) は、消費者がカテゴリー化を行う利点として、カテゴリーベース処理以外に、認知的努力を最小化し、取得する情報の価値を最大化する認知的経済性がある事 (Rosch, 1975) と、社会のコミュニケーションを円滑に行う役割の 2 点を挙げている。同様に高橋 (2011) は、カテゴリー化の情報処理により消費者相互の共通事項となるカテゴリーが創造され、カテゴリーが社会性を持つと指摘している。消費者の内面におけるカテゴリーが、同じ社会に

において共通化する傾向を持つならば、消費者が内面に持つ製品カテゴリーの類似評価の異質性は小さくなると考えることができる。

2.2.2 カテゴリー概念とカテゴリーの類似関係

カテゴリー構造の概念として、分類学的なカテゴリー構造、グレード化されたカテゴリー構造、アドホック・カテゴリー構造という3つの捉え方がある（新倉, 2005）。分類学的なカテゴリー構造では、動植物の系統分類のように、定義的特性によって分類が規定される。定義により明確な分類の境界が存在し、階層的な分類に適している。グレード化されたカテゴリー構造による分類では、代表的で典型的な分類対象をカテゴリーの中心と位置づけて、その中心メンバーからの近接関係による分類を行う。このため、明確な境界線は存在せず、カテゴリーの代表と認識される典型的対象からの類似性によって、各カテゴリーへの所属の程度が評価される。アドホック・カテゴリー構造は、消費者の特定の行動や目的のために、予め記憶にある情報や、その状況に直面して構造化される構造である。特定の状況における短期的なカテゴリー構造といえるが、同じ状況に接する頻度が増えると、次第に構造が記憶され、グレード化されていく知識構造である。

これらの3つカテゴリー構造の中で、定義的カテゴリーは個人毎に形成される性質のカテゴリー概念ではないし、アドホック・カテゴリーは定義より消費者の製品理解のために繰返し利用される知識では無い。このため、消費者がカテゴリーベース処理に用いるカテゴリー概念として、グレード化されたカテゴリーが関心の対象となる（Rosch, 1975; Tversky, 1977）。Rosch(1975) は実験により人が典型的なカテゴリーの対象（プロトタイプ）を基準にカテゴリー化している事を明らかにした。実験として、色と数字と線の図形の3種類の集合において人が類似を認知するさまを測定し、「赤っぽい色は真紅に近い」と知覚され、「11,52,103 といった数字は 10,50,100 に近い」と知覚され、「四角っぽい図形は真四角に近い」と知覚され易いことから、それぞれの集合において、真紅、10,50,100、真四角といった、プロトタイプがある事を示した。Tversky(1977) は、Rosch(1975) の結果を踏まえて集合の要素 a と b の類似 (similarity) を、集合 A と B をそれぞれの特徴の集合として、次式で表す対比モデル (Contrast Model) を示した。

$$S(a, b) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A) \quad (2.1)$$

ただし、係数 $\theta, \alpha, \beta \geq 0$ 、関数 f は集合の包含関係について単調なスカラー値を与える等の条件を満たす関数とする。関数 f は空集合に対して最小値となり、全集合に対して最大値をとる関数で、例として、集合の個数を返す関数がある。評価式 (2.1 式) は、統計指標としての類似係数 (similarity coefficient) に期待される単調性などの性質 (Janson & Vegelius, 1981; Baulieu, 1989) と同じ特性を仮定し、同種の算式で表現されている。対称性 $S(a, b) = S(b, a)$ については、Rosch (1975) の「赤っぽい色は真紅に近い」という知覚は「真紅は赤っぽい色に近い」という知覚より強いという結果を考察して、関数 f が満たす条件としていない。また、係数 $\theta, \alpha, \beta \geq 0$ の自由度を設けている点と、集合に対する関数としている点で、統計指標としての類似係数より一般的な表現となっているので、Tversky の対比モデルを使う際には、係数の値と関数 f の算式を定める必要がある。

Tversky(1977) の対比モデルは、その後のグレード化されたカテゴリー研究の端緒となり、単純な対比モデルを批判的に精緻化したり、分類対象の特徴に具体的な構造を与えるといった様々な研究が行われている (Ratneshwar & Shocker, 1991; Larkey & Markman, 2005)。また、市場に新たに投入された製品が、その製品が類似する製品のシェアを奪う類似性効果 (Tversky, 1972) についての研究も多い。これらの一連の研究では、類似評価の形成要因となるカテゴリー知識についての捉え方は一様ではなく、Tversky の対比モデルの類似評価を構成するための知識とする狭義の定義のほかに、消費者の記憶にあるカテゴリーの集合について知っている情報の集合 (知識) を合せてカテゴリーと捉える広義の定義もある (高橋, 2011)。本論では、消費者の行動面から製品カテゴリーの類似関係を推定する事に関心があり、カテゴリー知識の範囲の設定や、その形成プロセスは研究の範囲外とする。

カテゴリーベース処理に基づいた製品の理解による製品選択のモデルとしては、カテゴリー構造がネットワーク構造を持ち、消費者が評価対象をカテゴリーに照らし合わせようとしながら、その注意が特定のカテゴリーにフォーカスされながらネットワーク上を拡散するという、活性化拡散モデル (Collins & Loftus, 1975) がある。同様な概念として、ネットワークモデル (Aaker, 1996; Aaker, 2004) がある。オンラインショップの「この商品を買った人はこの商品も買っています。」という推奨は、このモデルの応用と解釈できる。消費者の関心は、実際に購入した製品から類似関係の高い他の製品に移り易いため、消費者の考慮集合 (Shocker, Ben-Akiva, Boccara & Nedungadi, 1991) に類似関係の高い製品は加わり易いと考えられる。

2.2.3 競争市場構造分析

市場における製品の競合関係と消費者の選好を同時に捉える競争市場構造分析は、製品カテゴリーと消費者セグメントを同時に対応付けて推定する手法と言える。井上(1992)は、競争市場構造分析の類型として、表2.2の通りデータの特徴と分析の表現のタイプで4類型を示し、競争市場構造分析は顧客サイドから、競争状態を把握することにより、マーケティング戦略の形成に関して有用な情報を提供しようとするものであると、位置付けている。

表 2-2 . 競争市場構造分析手法・モデルの類型

	空間的表現	非空間的表現
判断データ	製品マップなど、因子分析、多次元尺度構成法	Urban, Johnson & Hauser(1984) など
行動データ	Kamakura & Russell(1989) など	井上・中西(1990)、井上(1992) Grover & Srinivasan(1987,1989)、 Moe & Fader(2001)

井上(1992) 表1 を基にして作成。

顧客サイドから競争状態を把握する手法の一つに、直接観測できない消費者や製品カテゴリーの分類を、消費者の判断データや行動データから推定する潜在クラスモデル(Grover & Srinivasan 1987,1989; Moe & Fader 2001)がある。ただし、その一般形は常に識別可能と限らず(Goodman, 1974)、井上(1992)は識別性を保証するための条件としてカテゴリー構造を仮定する手法(MIGHTシステム)を提案している。また、潜在クラスモデルは離散的な分類ラベルを推定する手法なので、一般に表2.2の通り非空間的表現に分類される。行動データを用いた空間的表現の研究であるKamakura & Russell(1989)は、ブランドのシェアの変化の要因として、競合ブランドに対する競争力と、競合ブランドのプロモーションからの脆弱性の2軸の空間にブランドを配置する手法を提案している。

行動データを用いた競争市場構造分析の手法としては、マーケットシェアやブランドスイッチにおける競争関係に着目するアイデアがあるが、書籍や音楽CD市場のようにリピート購買が期待できない市場では安定的なマーケットシェアが存在せず、これらの手法で競合関係を評価するのが困難である。また、購買の有無という2値データに対しては、因子分析などの比率尺度を対象とした分析手法を適用できない。ところが、類似係数を用いて行動面から消費者セグメントを評価すると、自然と行動面から推定した製品カテゴリーに対応した消費者セグメントとなり、潜在クラスモデルのような推定のプロセスを用いることなく、

競争市場構造分析を実現できる可能性がある。

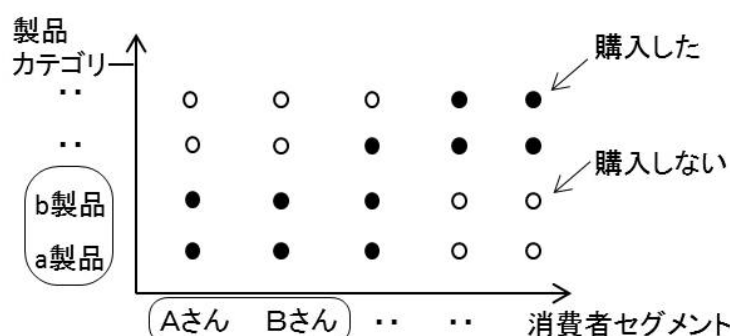


図 2-1. 製品カテゴリーと消費者セグメントの対応の概念

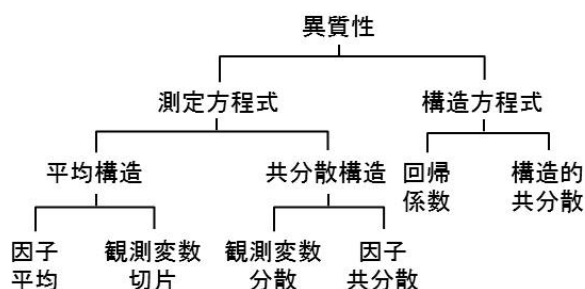
例えば、図 2-1 の通り、A さんが a 製品を購入し、B さんが b 製品を購入したとする。このとき、A さんと B さんが類似係数の評価として似ていれば、2 人の購入製品は共通している可能性が高いことを意味する。よって、A さんも b 製品を購入し、B さんも a 製品を購入している可能性が高い。さらに A さんに類似する人が N 人いた場合は、N 人が製品 a と b を同時に購入し易い事を表している。このため、製品 a と b を共通して購入し易い人が N 人いることになり、製品 a と b は類似係数による評価として類似していると推定できる。このように、消費者の類似を行動面で評価することは、競争市場構造分析の方針である製品カテゴリーに対応した消費者のカテゴリーを推定していると期待できる。以上の直感的な思考の真偽は、実証において検証すべき事項なので、非空間的な表現であるクラスター分析を用いて検証し、また同様に空間的表現においても検証する課題と考える。

2.2.4 消費者による製品選択の異質性の所在

ピースミール処理では、スポーツカーは若い男性に好まれるといったように、製品属性（スポーツカー）と消費者属性（若い男性）を対応させ、消費者属性によって消費者の異質性を説明することが多い。一方、カテゴリーベース処理の場合、性別や年齢でスポーツカーというカテゴリーの捉え方に差異があるのか議論の余地がある。そこで、カテゴリーベース処理に着目した消費者の製品選択の研究では、消費者によって異なる製品選択の異質性の所在を示すことが重要と考える。

カテゴリー知識を広義に捉える高橋（2011）は、カテゴリー分類に加え、製品に期待される属性を典型性としてカテゴリー知識の範囲に含め、その評価に消費者の異質性を置いて

いる。好まれ易い製品とは、理想的な属性を備えた典型的な製品に近い製品と考えられるので (Meyer-Levy & Tybout, 1989)、高橋 (2011) はカテゴリー知識と製品選択とを関連させて評価していると言える。これに対し、カテゴリーベース処理を製品の類似評価で捉える Tversky(1977) の対比モデルは、類似評価式に消費者の選好が含まれておらず、製品選択とカテゴリーベース処理を関連させていないモデルと言える。例えば A さんと B さんの車のカテゴリー知識と、2 人の車の選好とは独立であるという設定が可能なモデルである。このように、消費者のカテゴリー知識を選好と独立させて捉える考え方は、新倉 (2005) と高橋 (2011) の指摘における、消費者相互の共通事項となるカテゴリーが創造され、カテゴリーは社会性を持ち、社会のコミュニケーションを円滑に行う役割があるという指摘に整合している。



Ansari et al.(2000(2)), Fig.1 Sources of Heterogeneity を元に作成。

図 2-2. 共分散構造分析における異質性の所在

消費者の異質な選好をモデル化する類型として、Kamakura, Kim & Lee (1996) は選好の異質性と構造的異質性の 2 類型を示している。Ansari, Jedidi & Jagpal(2000(2)) は構造方程式モデルを念頭に、異質所の所在を測定方程式と構造方程式とさらに下位の構造に分類している (図 2-2)。Kamakura et al.(1996) の 2 類型は、Ansari et al.(2000(2)) の測定方程式と構造方程式に対応している。また、Ansari & Jedidi(2002) はベイズ統計を用いて異質性を扱う因子分析モデルを類型化して示している。因子平均に異質性を与えるモデルとして、西尾・竹内 (2011) は消費者の健康と環境に対する価値感調査を行い、消費者のデモグラフィックや情報の接し方や観測時期によって因子平均が異なる確認的因子分析モデルを示している。これに対して、通常の探索的因子分析は、異質性をモデル化しない分析法に類型化される。例えば、消費者毎に因子得点が異なる分析結果を得たとしても、性別や年齢などの属性によって因子得点の平均が異なるというモデル化を行っているわけではなく、それぞれの因子得点は今回の観測値に過ぎない。しかし、対応分析や因子分析によるプロダクト

マップの描画を行う分析では、消費者の選好がモデル化されていなくても、因子得点によって定まる布置に意味を見出そうとするのが通例である。全ての消費者は原点を中心とする分布に従うとは考えずに、消費者毎に特有の布置があると仮定して考察する。

因子分析を用いた空間的表現では、測定方程式の共分散構造から評価軸を構成し、平均構造に座標を対応させるのが通例である。このため、消費者のグループによって因子負荷量や因子を構成する観測変数の構成が異なる多母集団分析を行った場合、グループ毎に評価軸の構成が異なり、それぞれのグループの表現空間が異なり、双方のポジションを比較することが難しくなる。下表 2-3 に、以上のレビューを整理した。本研究で製品カテゴリーの類似関係が評価軸となる点は、研究課題 3 において後に議論する。

表 2-3. モデルにおける異質性の所在

	異質性の所在	異質性なし	異質性あり
製品カテゴリーと製品選択の関係	製品カテゴリーの類似関係	製品カテゴリーの類似関係は消費者に共通	製品カテゴリーの類似関係は消費者により異なる 消費者毎に異なるポジショニング
空間的表現	軸の構成	全消費者に 1 枚のポジショニングマップ、消費者の配置が異質	マップ、評価軸や製品の座標配置が異質
因子分析	因子負荷量	一般の因子分析、因子負荷量が共通で、因子得点が消費者で異質	因子負荷量が異質 Ansari & Jedidi(2002)
共分散構造分析	因子構成と付加量	一般の共分散構造分析	グループで因子構成が異なる 多母集団分析

2.2.5 製品カテゴリーの類似関係の 2 つの推定経路：知覚と行動

Tversky (1977) の対比モデルは、消費者が製品の特徴を知覚することでカテゴリー知識を形成すると考え、カテゴリー知識の形成要因から製品カテゴリーの類似関係を推定する方法である。これに対して、図 2-3 の通り、コンピューターサイエンスの学術的領域における協調フィルタリング Goldberg et al.(1992) は、POS データなどの購買履歴データを用いて行動面から製品のカテゴリーの類似関係を推定していると解釈できる。ただし、マーケティングとは異なる領域の手法である協調フィルタリングの研究において、Tversky (1977) の対比モデルの引用は見当たらず、行動面から製品カテゴリーの類似関係を推定できると示した研究も見当たらない。

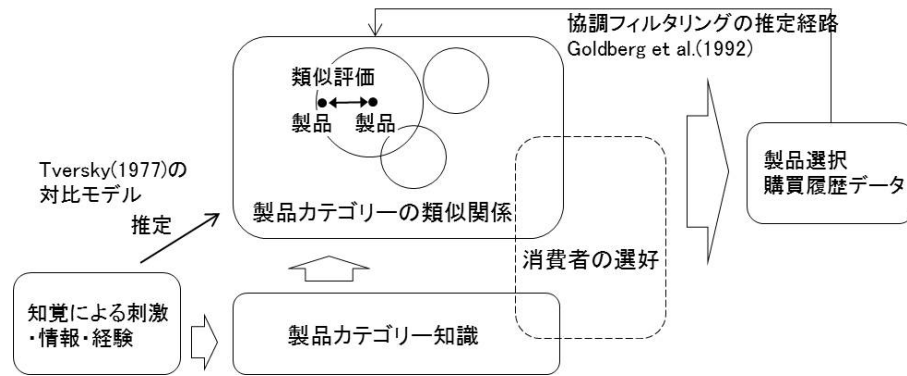


図 1-1 に加筆して再掲

図 2-3. 製品カテゴリーの類似関係の 2 つの推定法

協調フィルタリングは、オンラインショップなどのレコメンドシステムとして製品推奨に利用されるアルゴリズムで、製品間や消費者間の類似評価を利用して、消費者の製品選択を評価する仕組みである。レコメンドシステムのサーベイ研究である、Adomavicius & Tuzhilin (2005) と神鷲 (2007, 2008(1), 2008(2)) は、コンピューターサイエンスの観点から様々なレコメンド手法を取り挙げ、歴史的に最初の協調フィルタリングは、米国のゼロックス研究所の Tapestry システム (Goldberg et al., 1992) と位置付けている。この研究が行われた当時は、計算機のネットワーク化が進んで多数のニュースグループが登場し、ユーザーが処理しきれない膨大な情報が配信され始めていた。そこで、従来の属性に基づくフィルタリング法の限界が指摘され、各ユーザーが自身に有益な情報をシステム的に取捨選択する新しい情報フィルタリングの技術が求められていた。

今日、協調フィルタリングを用いたシステムが普及している背景には、オンラインショップの台頭がある。オンラインショップでは製品の陳列スペースに物理的な制約が無いため、売り手がインタラクティブに大量の製品や情報を提供しがちである。このため、消費者の情報探索コストや情報処理量の増加を招き、消費者にとって購買の困難性や不安としての知覚リスクが増す懸念がある。山本 (1999) は、知覚リスクの構造として不確実性と結果の重大性を挙げ、外部情報の探索行動が知覚リスクを低減する効果があると述べている。レコメンドシステムには、各消費者が考慮集合を作成する探索コストを低減し、購入候補の属性を比較するのを支援する役割がある (Häubl & Trifts, 2000)。また、製品推奨が消費者の購買に与える影響など (Lurie, 2004, Swaminathan, 2003)、マーケティングの研究分野では、アルゴリズムよりも消費者心理に対する効果に研究の関心が向けられている。

協調フィルタリングの有用性は、レコメンドシステムの今日の普及の現状から伺い知るこ

とができる。しかし、消費者の情報処理を踏まえたアルゴリズムではないために、消費者の異質性の所在の設定が議論されておらず、消費者の製品選択を評価する手法として課題がある。また、定量的なモデルとしても示されていない。このため、本研究では、協調フィルタリングに関する議論を踏まえつつ、行動面から製品カテゴリーの類似関係を推定する定量的モデルを構築することを目指す。

2.2.6 製品カテゴリーの類似関係を用いた製品のクラスター分析

Tversky の対比モデルの類似 (similarity) の評価式 (2.1 式) は、統計指標としての類似係数 (similarity coefficient) の一般的表現と解釈できる。類似係数は、2 値のプロファイルデータの近接関係 (proximity relation) を評価する指標に使われる指標であり、プロファイルデータは、 $n \times m$ の長方形列のデータで、主成分分析や因子分析の対象とされるデータ形式である (渡辺, 2008)。一般に、プロファイルデータの行と列に対応する対象は、 n 人の被験者の m 個の属性の評価値であったり、一人の被験者の n 個の刺激に対する m 個の属性の測定値であったりと様々であるが、本研究では、行が消費者、列が製品に対応するプロファイルデータを扱う。

表 2-4 の通り、多数の類似係数が提案されている。表 2-4 の計算式にある a, b, c, d 等の記号は、2 人の消費者あるいは 2 個の製品に関するクロス集計の集計値で、例えば製品 i と製品 j の類似関係を評価する場合は、表 2-5 の通りのクロス集計により、これらの値を求める。製品に代えて消費者間の類似関係を評価する場合は、クロス集計も同様に製品と消費者を入れ替えたものを用いる。

Wedel & Kamakula(2000) は、消費者セグメントのクラスター分析に用いる指標として 12 個の類似係数を紹介し、Albatineh, Niewiadomska-Bugaj, & Mihalko(2006) は 22 個の類似係数を紹介している。しかし、クラスター分析の際にどの類似係数を選択するかという基準は、殆ど示されていない (Janson & Vegelius, 1981; Baulieu, 1989; Duarte, Santos, & Melo, 1999)。また、交互作用統計量も類似係数のひとつとみなせるが、交互作用統計量を類似係数として扱う研究は見当たらない。

類似係数は、離散的な属性を持つ対象の判別に有益であり、ゲノム解析 (Duarte. et al.,1999) やレコメンドシステム (芳賀, 2008) やテキストマイニング (Shawe-Taylor & Cristianini, 2004) の分析データとして用いられる。また、多次元尺度構成法を用いた選好分析にも利用

表 2-4 . 主な類似係数の算式の一覧

No.	名称	類似係数の算式
1	交互作用統計量	$\frac{ad-bc}{\sqrt{R_1 R_2 C_1 C_2 / n}}$ (*1)
2	Jaccard	$a / (a + b + c)$
3	Czesanowski	$2a / (2a + b + c)$
4	Ochiai	$a / \sqrt{(a + b)(a + c)}$
5	Cohen's Kappa	$\frac{2(ad-bc)}{2(ad-bc) + n(b+c)}$
6	Sokal and Sneath 2	$a / (a + 2b + 2c)$
7	Russel and Rao	a / n
8	Yule's Q	$(ad - bc) / (ad + bc)$
9	Kulczynski	$a / (b + c)$
10	Rogers and Tanimoto	$(a + d) / (a + b + 2c + 2d)$
11	Sokal and Sneath 1	$2(a + d) / (2a + 2d + b + c)$
12	Simple matching	$(a + d) / n$
13	Hamann	$(a + d - b - c) / n$
14	主効果の相乗平均	$\sqrt{(a + b)(a + c)} / n$ (*2)

Wedel & Kamakura(2000) 表 5.1 を基にして作成。

*1. 交互作用統計量を二乗すれば独立性の検定に用いるカイ二乗統計量となる。

本研究では Yates の修正を用いる。 $R_1 = a + b$, $R_2 = c + d$, $C_1 = a + c$, $C_2 = b + d$

*2. 類似係数では無いが比較対象として表に記載した。

表 2-2. 製品 i と製品 j の類似係数を計算するためのクロス集計

		製品 j		
製品 i		×	計	
		$a_{i,j}$	$b_{i,j}$	$R_{i,1}$
	×	$c_{i,j}$	$d_{i,j}$	$R_{i,2}$
	計	$C_{j,1}$	$C_{j,2}$	n

$a_{i,j}$ = 製品 i と j を共に購入した消費者数。

$b_{i,j}$ = 製品 i を購入し、製品 j を未購入の消費者数。

$c_{i,j}$ = 製品 j を購入し、製品 i を未購入の消費者数。

$d_{i,j}$ = 製品 i と j を共に未購入の消費者数。

される (片平, 1987)。

製品カテゴリーの類似関係を類似係数で表現した後、離散的な分類ラベルを与える手法としてクラスター分析がある。Punj & Stewart(1983) は、マーケティングにおけるクラスター分析のサーベイ研究として、どのような分析手法や評価基準が採用されているかを示している。分析手法には、階層的分類と非階層的分類の大きく 2 つがある。前者の階層的分

類法には、ウォード法、群平均法、最長距離法、最短距離法の4つの代表的な手法があり、これらは最も近い2つの分類対象の凝集を逐次繰り返し、新しい分類を作成する分析手法である。それぞれ、新しく併合した分類と他の分類との距離を決めるパラメーターの設定に違いがあり、Lance & Williams (1967) と DuBien & Warde (1979) は、このパラメーターの差異がクラスターの階層構造における空間の歪みの程度に対応する事を示している。非階層的分類法としては、k-means 法が広く知られているが、直交座標系のデータを対象とする手法であるため、距離の三角不等式を満たす保証のない類似係数には適用できない。この課題に対応し、Kaufman & Rousseeuw (1990) は k-means 法のアイデアを踏まえた非階層的分類法として、PAM (Partitioning Around Medoids) 法を提案している。PAM 法は、類似係数に適用できる非階層的分類法である。

これらのクラスター分析は、教師なし分類であるため、分類結果の妥当性評価や、適切なクラスター数を導く事が課題となる。クラスター分析の妥当性を評価する手法として、再現性・均等性・外的基準・内的基準の4つの定量的基準と、定性的な解釈可能性の基準により、総合的に評価を行うクラスターバリデーション (Milligan, 1996; Gordon, 1999) がある。階層的クラスター法では、最初に作成するグループが次に作成するグループに影響を与え、クラスターの初期の振る舞いの偶然が、最終的な分類結果に大きな影響を与える懸念がある。また、凝集の繰り返しが空間の歪みを引き起こし、空間膨張的あるいは空間収縮的と呼ばれる好ましくない分類結果を導くことが知られている (Yaduhisa, Takeuchi, & Inada 1999)。このため、クラスターバリデーションでは、分類対象を乱数で抽出して分類の再現性を確認するといった、交差確認法が重要となる (Gordon, 1999)。Punj & Stewart (1983) のサーベイ研究からも、分類の再現性を評価基準とする研究が多い事が判る。

2.2.7 消費者の選好構造の空間的表現

一般に類似係数は距離の三角不等式を満たさないため、ユークリッド距離空間を前提とする統計解析の手法が利用できないという課題がある。このため Tversky の対比モデルの類似評価に対して行える分析手法は、クラスター分析など一部の統計分析手法に限定され、選択確率や消費者の類似の分布といった、連続的な比率尺度としての結果を導く事ができない。

これに対して、空間的な表現は、連続的な近接関係の情報を保持しているため、消費者の選好構造を空間的に表現できれば、さらに様々な分析手法を適用することができる利便性が

ある。この場合の空間的表現としては、2次元への射影による視覚化ではなく、対象をユークリッド距離空間に埋め込み、様々な統計手法を適用可能にする事を意図している。マーケティングにおける空間表現としての布置には、製品間の関係を表す目的と、消費者間の選好の関係を表す2つの目的がある（阿部, 2003）。これらの布置を用いると、製品や消費者のポジションを示すことができるため、製品間の競合や消費者セグメントを解釈するのに有益である。

消費者の製品に対する選好を空間上の距離に対応させる手法として、ベクトルの方向で捉える理想ベクトルモデル（狭義のバイプロット）と、2点の距離で捉える理想点モデルの、大きく2つがある（阿部, 2003; 柳井他, 1990）。因子分析を用いて因子スコアを点として、因子パターンをベクトルとして描くバイプロットは、ベクトルモデルに分類される。後者の理想点を描く手法には多次元展開法があるが、解の推定が難しい手法とされる（前川, 1999）。両者のモデルを比較すると、消費者の製品に対する選好を距離で表す理想点モデルの方が直感的に扱いやすいが、購買量の多さといった選好の強さを表現するのが難しい。理想ベクトルモデルの場合は、ベクトルの長さで選好の強さを表し、ベクトル間の角度で選好の差異を表すことができる。阿部（2003）は理想ベクトルモデルと理想点モデルなどの複数のジョイントスペースを比較し、空間的表現としての優劣を評価する3つの基準として、（1）空間的表現を与えるモデルの適合度（対数尤度）、（2）モデルによる製品選択の予測確率、（3）ヒット率（実際の製品選択の正解率）の3点を挙げている。

製品カテゴリーの類似関係を空間的に表現する場合、距離の三角不等式を満たさない類似係数行列を直交座標系に埋め込むモデルとして、多次元尺度構成法（渡辺, 2008）がある。また、類似係数行列が内積とみなせる場合には、カーネル法（Shawe-Taylor & Cristianini, 2004）により高次元の空間に埋め込むことができる。

2.3 本研究における主要課題

本節では、先行研究のレビューを踏まえて研究課題を明らかにする。前節における先行研究のレビューより、製品カテゴリーの類似関係を、消費者の製品に対する知覚内容から推定するTversky(1977)の対比モデルを端緒とする多数の研究があることが分かった。消費者の知覚を用いるためには、個々の消費者の知覚に関するデータを集める必要があるが、実務的な応用を考慮すると、頻繁に多くの消費者に対してアンケート調査を行う事は、コストの面

で課題がある。消費者の知覚データを用いずに、購買履歴データから消費者の購買を推定する Goldberg et al.(1992) の提唱した協調フィルタリングは、Tversky(1977) の対比モデルと同じく製品間や消費者間の類似評価を用いる手法であるが、コンピュータサイエンスの学術分野におけるアルゴリズムであり、カテゴリー概念を踏まえたモデルでは無いため、行動面から製品カテゴリーの類似関係を推定する妥当性は示されていない。行動面のデータを用いる利点として、小売店のPOSデータやオンラインショップの販売履歴データを利用し、データ収集のための追加コストを払うことなく、既存の流通・販売システムと一体となったりリアルタイムの市場分析を実現できる事がある。行動面から製品カテゴリーの類似関係を推定する手法は、高い利便性が期待できるものの、推定手法と推定結果の妥当性を検証する必要がある事が課題となる。

次に、消費者のカテゴリーベース処理について、消費者の製品選択に影響を与えるカテゴリー知識が製品選択に及ぼす影響に関する研究として、同じ買い物カゴに同時に投入され易い製品カテゴリーの組み合わせに関する研究 (Manchanda, Ansari & Gupta, 1999) や、共変信念の影響に関する新倉 (1997) の研究などの特定の要因に関する精緻な研究はあるものの、カテゴリー知識によって構成される製品カテゴリーの類似関係と過去の購買履歴を用いて、将来の製品選択を説明する一般的なモデルが示されていない事が分かった。協調フィルタリングのアルゴリズムは、ひとつの解釈として行動面からカテゴリーベース処理を捉えていると考えることもできようが、消費者の異質性の所在などが明示されておらず、マーケティングのモデルとしての十分な要件を満たしていない。そこで、製品カテゴリーの類似関係と製品選択とを関係付けるモデルを提示する事は、消費者の行うカテゴリー化が製品選択に与える影響を明らかにする枠組みとして有用と考える。

最後に、消費者の製品選択における情報処理プロセスの研究において、ピースミール処理には多属性態度モデルという一般的な分析のフレームがあり、多数の研究が行われているのに対し、カテゴリーベース処理では一般的な分析のフレームが提示されていないという課題がある事が分かった。そのような分析のフレームが提示されていない理由として、製品カテゴリーの類似関係の評価式である類似係数が、一般に距離の三角不等式を満たさないため、分析対象を線形空間として表現しがたく、分析に利用できる統計手法が限られる事が挙げられる。

以上の考察を踏まえ、本研究における研究課題として、(1) 消費者の行動面から製品カテゴリーの類似関係を推定する手法の提示、(2) 製品カテゴリーの類似関係と製品選択との関

係を示すモデルの提示、(3) 競争市場構造分析の空間的表現として多様な消費者の選好構造を捉え、消費者の選好構造と売上との関係について新たな知見を示す、以上の3点を研究課題とする。

消費者の行動面から製品カテゴリーの類似関係を推定する手法の提示

先行研究のレビューにより、消費者のグレード化された製品カテゴリーの構造を、類似関係として定量的に推定するモデルとして、Tversky(1977)の対比モデルを端緒とするさまざまな研究がある事が分かった。これらの手法は、消費者が知覚する何らかの製品の特徴が、カテゴリー知識の形成の刺激となっていると考え、2製品の特徴の差異の類似を評価するものである。一方で、オンラインショップのレコメンドシステムのアルゴリズムである協調フィルタリングは、2製品の購入者が共通し易い程度を類似関係として評価して製品推奨に利用するため、行動面から製品の類似評価をしていると解釈できる。

消費者の内面にあって直接観測できない製品カテゴリーの類似関係を推定する手法として、カテゴリーの形成要因である消費者の知覚から推定する手法と、カテゴリーベース処理を反映した行動結果から推定する2つの手法は、どちらかが優れているという関係ではなく、補完的な関係にあると考えられる。知覚から推定する手法は、製品の特徴やその構造の一覧を説明変数として仮定するため検証的であり、行動結果から推定する手法は、説明変数を設けないという点で記述的である。双方の手法があれば、記述的手法から探索的に知見を見出し、その知見から検証するモデルを構成するサイクルが成立する。また、製品の特徴に対する知覚データを入手するには質問調査を行う手間とコストが必要となるが、行動面のデータは購買時のPOSデータやオンラインショップの購買履歴を利用できるため、POSシステムを導入している販売店やオンラインショップにとっては、既存のシステムにあるデータを用いてリアルタイムで市場分析を行える利便性がある。

製品カテゴリーの類似関係を行動面から推定する手法としては、コンピュータサイエンスのアルゴリズムとしての協調フィルタリングがあるが、マーケティングの手法として検証されているとは言いがたい。そこで、第1の研究課題として、Tversky(1977)の対比モデル同様に、2製品に対する消費者の購買の独立性を類似係数で評価し、この類似データに対してクラスター分析を適用し、分析結果の製品カテゴリーの妥当性を評価する(図2-4)。マーケティング研究におけるクラスター分析の利用についてサーベイ研究した Punj & Stewart(1983,

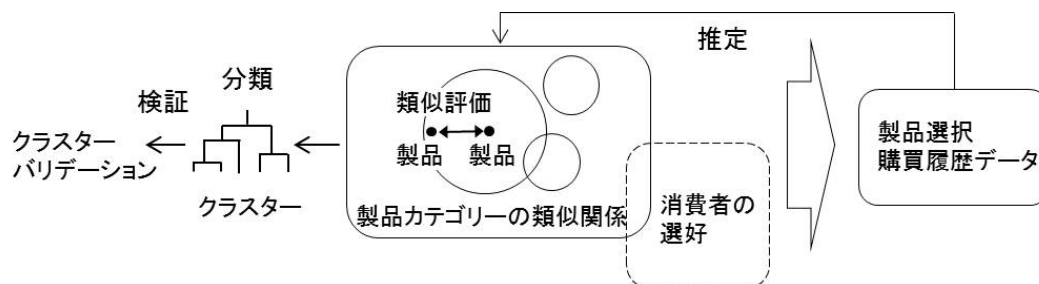


図 2-4. 行動面から製品カテゴリーの類似関係を推定する妥当性の検証

表 3) によると、ランドの修正係数による分類結果の再現性がクラスター分析の妥当性の評価として重視されている。また、統計分析の観点から Gordon(1999) は、クラスター分析は明確な評価基準が無い教師なし分類であるため、類似係数とクラスター分析の選択、および評価基準について網羅的に慎重に行う方針を提案している。本研究では、これらの先行研究が示す手順に従い、特に製品カテゴリーに関する既存の知見との整合性として、外的基準や解釈可能性を重視し、行動面から製品カテゴリーの類似関係を推定する手法の妥当性を検証する。

製品カテゴリーの類似関係と製品選択との関連を示すモデルの提示

消費者の製品選択におけるピースミール処理とカテゴリーベース処理の2つの情報処理プロセスにおいて、ピースミール処理には製品属性から購買を説明する多属性モデルがあるのに対し、カテゴリーベース処理には同様の一般的なモデルが示されていない事が分かった(図 2-5)。行動面の類似から消費者に推奨を行う協調フィルタリングのアルゴリズムは、消費者のカテゴリー化の情報処理プロセスを踏まえたロジックと解釈できるものの、消費者の異質性の所在が明示されておらず、マーケティングのモデルとしての十分な要件を満たしていない。そこで、コンピューターサイエンスの分野である協調フィルタリングのアルゴリズムのアイデアを、消費者の購買行動に沿ってモデルとして示すことができれば、同じテーマを扱う2つの学際的研究になると期待できる。

モデルの構成として、説明変数に製品カテゴリーの定量的表現である類似係数を用い、消費者の製品選択を説明するモデルを提示する事を目指す。消費者の内面にあるカテゴリーの表現として、Tversky(1977)の対比モデルの通り製品間の類似係数で表現する立場と、製品の特徴の構造を含めて広義に捉える立場がある事を踏まえると、狭義の類似係数のみで製品

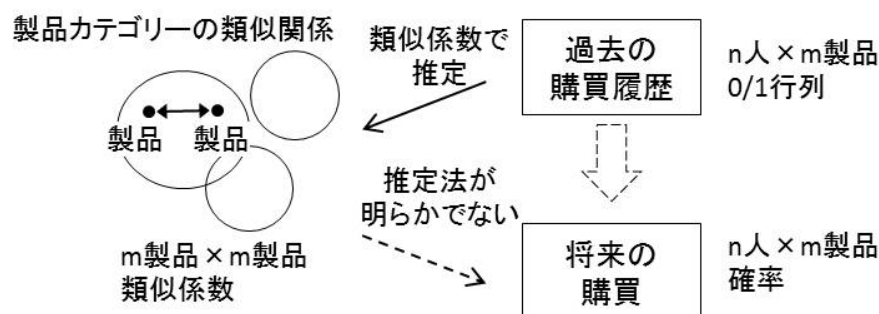


図 2-5. 製品カテゴリーの類似関係を用いた製品選択予測

選択を説明できるか否かの検証は、対比モデルの有用性を議論する際の一つの評価材料になると考える。実証においては、データの入手に関する現実的な制約から、POSデータを用いて行動面から推定した製品カテゴリーの類似関係を用いたモデルの検証を行う。また、モデルにおける消費者の異質性の所在については、異質性の扱いに関する先行研究を踏まえ、異質性の扱いが少ない単純なモデルを選択する方針で望む。単純なモデルは、実証において解釈が容易であり、モデルの限界の発見を通して新たな知見を見出すのに都合が良いと考える。

消費者の選好構造の空間的表現

先行研究のレビューの結果、消費者のカテゴリーベース処理に着目した分析の有用性に関する実証的な議論が少ないことが分かった。その理由として、前述の通り製品カテゴリーの類似係数と製品選択の関係を捉える一般的なモデルが提示されていない事がある。また、類似係数が一般に距離の三角不等式を満たさないため、分析手法の制約が大きい事が挙げられる。例えば、クラスター分析は類似係数に対して利用可能な統計手法であるが、推定結果として離散的な分類ラベルを求める分析手法であるため、特定の製品の潜在顧客の分布を評価したり、その連続的な変化を評価するのに適していない。

そこで、第3の研究課題として、消費者の製品カテゴリーに対する選好構造を線形空間上に表現し、消費者の選好構造と売上との関係について新たな知見を示す。ここで消費者の選好構造とは、製品からの距離によって選好の強さが関連付けられた消費者の空間的な分布とする(図 2-6)。空間的に表現された選好構造に対しては、線形空間上のデータに対して適用できる様々な統計手法を適用し、製品カテゴリーに着目した様々な分析が可能になる。例え

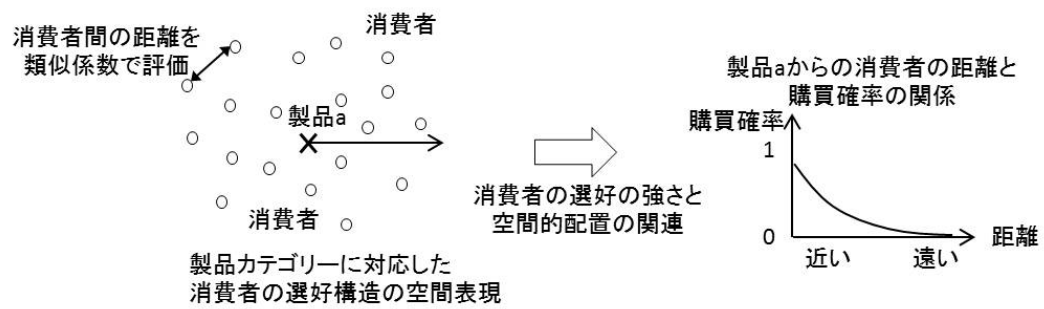


図 2-6. 競争市場構造分析としての空間的表現の評価

ば、タイムスタンプのあるPOSデータを利用して、行動面から製品カテゴリーの類似関係の連続的な変化を捉えたり、消費者の製品カテゴリーの類似関係の形成プロセスを評価したり、その結果を製品の売上推移と関連付けるといった分析が可能になると考える。

第3章 行動面からの製品カテゴリーの類似関係の推定

3.1 はじめに

第3章では、行動面から製品カテゴリーの類似関係を定量的に推定する手法を提示し、その妥当性を検証する。検証手法としては、候補となる類似評価式毎にクラスター分析を適用し、クラスターバリデーションの基準で相対的な優劣を評価する。Tverskyの対比モデルでは、統計分野において類似係数と呼称される指標群の一般式として類似評価式が与えられている事を踏まえ、主要な類似係数を類似評価式の候補とする。候補とする類似係数のリストは、消費者セグメントのための指標として広く知られている指標が含まれるように作成する。類似係数の優劣の判定において、クラスター分析の分類結果は教師なし分類となるため絶対的な評価基準が存在しない。そこで、多面的に相対的な評価を注意深く行う。ただし、単なる網羅的な評価とならないように、類似係数の優劣の評価軸として交互作用と主効果を設定し、望ましい類似評価の特徴について明らかにする。

一般に、連続変量の記述多変量解析には、変量の変動を特異値分析で近似要約する主成分分析および交互作用を近似するMandel(1971)の交互作用解析と共に、類似係数に基づくクラスター分析の2つの流れがある。一方、2値変量の記述多変量解析においてはコレスポンデンス分析に代表される特異値分析等の近似要約はよく用いられるものの、クラスター分析については、どのような類似係数がどのような分析結果を導くかについて殆ど研究されていない(Janson & Vegelius, 1981; Baulieu, 1989; Duarte et al., 1999)。そこで本研究では、2値変量の解析で提唱されている様々な類似係数について、クラスター分析の観点から好ましい結果を導く類似係数の特徴を実験的に示す。また、類似係数の交互作用統計量による偏導関数を求め、クラスター分析に適していない類似係数と交互作用統計量との関係の特徴を示す。

本研究の結果として示す通り、いくつかの類似係数は空間収縮を引き起こすなど必ずしも

好ましくないクラスター分析を導く。特にコレスポンデンス分析における距離の役割を果たすカイ二乗距離は、検定統計量に含まれる行列の交互作用の影響に比べて主効果情報の影響を強く受け、主効果が大きくなると距離が小さくなる特徴を持つため、結果として著しい空間凝縮を見せる。この事は、カイ二乗距離行列の低次元近似を原理とするコレスポンデンス分析においても、応用上の問題を引き起こす要因になると考えられ、大規模疎行列に対するコレスポンデンス分析は上位2～3の固有値を用いても、その累積寄与率は極めて低くなるのが通例である。従って、そのような行列に対する記述的な分析の第一段階として、類似係数に基づくクラスター分析を行う事は有用と考える。例えば、膨大な2値データを記述的に解析するゲノム解析 (Duarte, et al.,1999) やマーケティングのPOSデータ解析 (芳賀, 2008) の分野において、類似係数の選択が課題とされている。

本研究では、類似係数が交互作用と主効果の関数で構成されている事に着目し、この2軸に沿って評価の優劣を検証する。この2要因を個別に分類に用いるとすると、まず観測対象に固有の値である主効果は観測対象に全順序を与え、分類というより主効果のランキングを与える。一方、交互作用は2項関係なので分類に適切な指標と期待できる。そこで主効果の影響が少なく、交互作用との相関が高い類似係数ほど分類に適切な指標になると予想する。

この予想を検証するために、主効果と交互作用が無相関となる事を踏まえて、類似係数を交互作用から主効果へと相関の高い順に並べる。そして、この順序と分類の妥当性の評価が一致する事を検証する。この検証プロセスの課題は、分類の妥当性の評価を客観的に示す事である。教師あり分類の場合であれば、真の分類との一致を評価すれば良い。しかし本論では教師無し分類を評価するため、クラスターバリデーション (Milligan, 1996; Gordon, 1999,7章) の手順に従って、多面的な評価基準を用いる。具体的には再現性・均等性・外的基準・内的基準の4つの定量的基準と、定性的な解釈可能性の基準で評価する。

ところで、この検証には考察すべき課題がある。第1の課題は、交互作用の指標の選択である。本研究では、対応表の独立性の検定に用いるピアソンのカイ二乗統計量のルートに符号を付け総度数の平方値で割って $[-1, 1]$ の値域を持つように基準化した値を交互作用統計量とする (表 3-2 の M1)。二乗する前の符号を戻すのは、交互作用がプラスかマイナスかで分類結果が異なる事に配慮している。類似係数の Psi (表 3-2 の M7) は交互作用統計量に絶対値を付けた指標であるが、プラスの交互作用が存在するためにはマイナスの交互作用も必ず存在しており、符号が持つ情報を活かすかどうかで分類結果も異なる。また、ピアソンのカイ二乗統計量以外にも Goodman (1996) の連関モデルのように乗法的に交互作用を捉える

考えもあろう。より良い交互作用の指標の選択については、当初の予想が肯定的に検証できた後の課題とし、本研究では主効果による偶然一致の影響を除いた指標で、主効果の影響を受ける標準偏差で基準化した指標が、他の類似係数より妥当な分類を与える事のみを検証する。

第2の課題は、教師なし分類の検証に強力な基準が無い事である (Milligan, 1996; Gordon, 1999, 7章)。どのような分類結果に対しても、分析者は「これが私の意図した分類です。」と主張できる。このため、教師なし分類の検証では、複数の評価基準を併用する多面的評価が望ましい。また、個々の基準の評価についても、データに依存して偶然良い結果を得る可能性を無視できない。そこで本研究では、2-フォールド交差確認法を用いて分類を作成したり、検証指標の有意水準を評価する確率分布が分かっていない場合は乱数を用いたシミュレーションで p 値に代わる経験的パーセンタイル点を求めたりして、検証の信頼性を高める事とする。

本論では、次の2章で検証の手順と分類の妥当性を評価するクラスターバリデーションについて述べる。続く3章で評価基準毎に実証分析を行い、結果を整理して考察する。実証データには音楽CDの個人ID付き販売POSデータを使用する。そして最後に4章で結果を考察し、今後の課題を述べる。

3.2 検証の手法

3.2.1 類似係数

0と1を要素とする m 行 n 列の2値行列 $M = \{m_{ij}\}$ を、行方向が m 個の観測対象で、列方向が n 個の2値変量の属性に対応するデータセットとする。このとき観測対象 i_1 と i_2 の類似係数を、 i_1 と i_2 の n 個の属性値の 2×2 クロス集計表 (表 3-1) と、各類似係数の算式 (表 3-2) より計算する。

表 3-2 は Wadel & Kamakula (2000, p.46, 表 3-5.1) が最も重要とする類似係数のリストに、M1 交互作用統計量と M2 Cohen's Kappa (Cohen, 1960) と M12 カイ二乗距離と M16 主効果の相乗平均の4指標を追加して作成している。Albatineh et al. (2006) は、このほとんどの類似係数の出典をサーベイしている。しかしながら、これまで交互作用統計量は類似係数として扱われていない。また、主効果の相乗平均は類似係数では無いけれど、比較の基

表 3-1. 観測対象 i_1 と i_2 のクロス集計表

		観測対象 i_2		
観測対象 i_1		1	0	計
	1	a	b	r_1
	0	c	d	r_2
	計	c_1	c_2	n

ただし、集計値 a, b, c, d は、

$$a = \sum_j m_{i_1 j} \cdot m_{i_2 j},$$

$$b = \sum_j m_{i_1 j} \cdot (1 - m_{i_2 j}),$$

$$c = \sum_j (1 - m_{i_1 j}) \cdot m_{i_2 j},$$

$$d = \sum_j (1 - m_{i_1 j}) \cdot (1 - m_{i_2 j})$$

また行和と列和と総和を $r_1 = a + b$, $r_2 = c + d$,

$$c_1 = a + c, c_2 = b + d, n = a + b + c + d.$$

準とするために検証する類似係数のリストに加える。この中で良く知られた類似係数は M4 Jaccard と M13 Simple matching であろう。これらの類似係数行列は半正定値となるので、非類似係数には自然な距離が対応する (Gower & Legendre, 1986; Gordon, 1999) ことが知られている。また M2 Cohens's Kappa は見かけの一致率から偶然一致率を差し引いているので、交互作用に近い特徴を持っている。ただし標準偏差で基準化していない。M12 カイ二乗距離は表 3-1 の 2×2 クロス集計表から計算する指標では無いが、同じ 2 値行列 M を用いて計算する指標であり、分布の同等性という優れた特性 (大隈 他, 1994) を持っている事と、またコレスポンデンス分析に用いる距離として知られている事からリストに加えている。¹ ところで、表 3-2 の類似係数の記載順は、後述する実証データにおいて交互作用統計量と相関が高い順としている。また表 3-2 の非類似係数は、分類のための距離行列に相当する非類似行列を求めるための算式で、非負となるように (定数 - 類似係数) の式で定義している。

多くの類似係数の比較あるいは評価をしている先行研究として、Batagelj & Bren(1995) と Warrens(2008) は類似係数が分数の形式を持つために、分母が 0 に近い指標を定義できない問題を指摘し、修正式を提案している。Janson & Vegelius(1981) と Baulieu(1989) は類似係数が持つべき好ましい複数の性質、例えば対称性などを示し、各類似係数がそれらの性質を満たすか否かを確認している。Fleiss(1975) は類似係数の計算において 2×2 クロス集計

¹ オッズ比については、M9 $Yule's Q = \frac{\text{オッズ比}-1}{\text{オッズ比}+1}$ の関係により、Yule's Q と単調に一対一対応する。先行研究ではオッズ比より Yule's Q を採り上げる事が多い慣例にならない (Wadel & Kamakula, 2000; Warrens, 2008)、オッズ比を表 3-2 のリストから割愛している。

表 3-2. (非)類似係数の一覧

名称	類似係数 s_{ij}	非類似係数 d_{ij}
M1 交互作用統計量	$\frac{ad-bc}{\sqrt{r_1 r_2 c_1 c_2}}$ (*1)	$1 - \frac{ad-bc}{\sqrt{r_1 r_2 c_1 c_2}}$
M2 Cohen's Kappa	$\frac{2(ad-bc)}{2(ad-bc)+n(b+c)}$	$\frac{n(b+c)}{2(ad-bc)+n(b+c)}$
M3 Ochiai	$a/\sqrt{(a+b)(a+c)}$	$1 - a/\sqrt{(a+b)(a+c)}$
M4 Jaccard	$a/(a+b+c)$	$(b+c)/(a+b+c)$
M5 Sokal & Sneath 2	$a/(a+2b+2c)$	$(2b+2c)/(a+2b+2c)$
M6 Czesanowski	$2a/(2a+b+c)$	$(b+c)/(2a+b+c)$
M7 Psi	$\frac{ ad-bc }{\sqrt{r_1 r_2 c_1 c_2}}$	$1 - \frac{ ad-bc }{\sqrt{r_1 r_2 c_1 c_2}}$
M8 Kulczynski	$a/(b+c)$ (2*)	$\max_{i,j}\{s_{ij}\} - s_{ij}$
M9 Yule's Q	$(ad-bc)/(ad+bc)$	$2bc/(ad+bc)$
M10 Russel & Rao	a/n	$(b+c+d)/n$
M11 Rogers & Tanimoto	$(a+d)/(a+2b+2c+d)$	$2(b+c)/(a+2b+2c+d)$
M12 χ^2 metric	undefined	$\sum_{k=1}^n \frac{n}{m_{k\bullet}} \left(\frac{m_{ki}}{m_{\bullet i}} - \frac{m_{kj}}{m_{\bullet j}} \right)^2$ (*3)
M13 Simple matching	$(a+d)/n$	$(b+c)/n$
M14 Sokal & Sneath 1	$2(a+d)/(2a+2d+b+c)$	$(b+c)/(2a+2d+b+c)$
M15 Hamann	$(a+d-b-c)/n$	$(2b+2c)/n$
M16 主効果の相乗平均	$\sqrt{(a+b)(a+c)}/n$	$1 - \sqrt{(a+b)(a+c)}/n$

*1. a, b, c または d のいずれかの値が 5 以下の場合は、イエーツの修正を用いる。

*2. $b+c=0$ の場合は、 $a/(b+c)$ を a に置換える。

*3. $m_{i\bullet}$ は行列 $M = \{m_{ij}\}$ の i 行の行和、 $m_{\bullet j}$ は j 列の列和とする。

表の偶然一致を除く事の重要性を指摘し、偶然一致の確率分布について論じている。Duarte et al., (1999) は 8 個の類似係数で 27 個のサンプルをクラスター分析し、類似係数の順位相関とデンドログラムの類似について論じている。このように多数の類似係数を比較する研究はいくつかあるものの、類似係数毎の分類結果の妥当性を評価する研究は見当たらない。

次に、本論の予想に関わる議論として、類似係数の構成における交互作用と主効果の関係について考察する。表 3-1 の 2×2 クロス集計表は a, b, c, d の 4 つの観測度数情報を持っており、これは以下の 4 つのクロス集計表の統計量から一対一変換で表現できる同じ情報量を持つ。

$$\begin{array}{ll}
1 \text{ 総度数} & n = a + b + c + d \\
2 \text{ 行の主効果} & x_1 \stackrel{def}{=} (a + b)/n \\
3 \text{ 列の主効果} & x_2 \stackrel{def}{=} (a + c)/n \\
4 \text{ 交互作用統計量} & y \stackrel{def}{=} \frac{ad - bc}{\sqrt{r_1 r_2 c_1 c_2}}
\end{array}$$

すなわち、度数 a, b, c, d と統計量 n, x_1, x_2, y の間に次式の関係式が成り立つ。

$$\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = n \begin{pmatrix} x_1 x_2 \\ x_1(1 - x_2) \\ (1 - x_1)x_2 \\ (1 - x_1)(1 - x_2) \end{pmatrix} + ny \sqrt{x_1(1 - x_1)x_2(1 - x_2)} \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}$$

上式を表 3-2 の類似係数の算式に代入すれば、表 3-3 の通り類似係数を主効果と交互作用統計量の関数として記述する事ができる。この変換により、M12 カイ二乗距離の行和 $m_{k\bullet}$ を除いた全ての類似係数の分母分子で総度数 n がキャンセルされ、類似係数を 4 変数から 3 変数の関数に置き換えることができる。そして、大規模疎行列に類似係数を適用した場合の類似係数と主効果および交互作用統計量との関係を偏導関数を用いて評価できる利点がある。すなわち大規模疎行列では行と列の主効果 x_1, x_2 が小さくなる傾向を踏まえ、まず表 3-3 の関数式に $x_1 = x_2 \stackrel{def}{=} x$ を仮定して関数式を整理し (表 43-の第 2 列)、これを交互作用統計量 y で偏微分して $x = 0$ を代入することで主効果が小さい場合の類似係数と交互作用統計量との関係が評価できる。ここで、主効果が等しいという条件は簡易に計算を行うために設けた条件であるが、類似係数に期待される性質は主効果が等しい場合においても満たされるべきである。

結果は表 3-4 の通り、行と列の主効果が等しい ($x_1 = x_2$) という条件の下で、類似係数を 3 つのタイプに分類できた。最初のタイプ I と分類した類似係数は、M1 交互作用統計量と M2 Cohen's Kappa と M7 Psi で、いずれも主効果の影響を受けない。次のタイプ II と分類した類似係数は、M3 Ochiai、M4 Jaccard、M5 Sokal & Sneath 2、M6 Czesanowski、M8 Kulczynski の 5 個で、興味深いことに交互作用統計量と主効果による 2 つの影響の程度が等しく対称な $(1 - \text{主効果})(1 - \text{交互作用統計量})$ の関数として記述できる。その他のタイプ III は M9 Yules's Q 以下の 8 個の類似係数で、このうち 7 個は主効果 $x = 0$ の偏微分係数がゼロとなり、主効果がゼロに近づくと交互作用統計量に対して無相関となる。またカイ二乗距離については、分母に主効果 x が単独で残るため、主効果 x が小さくなると交互作用統計

表 3-3. 主効果と交互作用統計量と類似係数の関係式 - その 1

名称	類似係数の算式 (*1)
M1 交互作用統計量	y
M2 Cohen's Kappa	$2y \cdot \frac{\sqrt{x_1(1-x_1)x_2(1-x_2)}}{x_1+x_2-2x_1x_2}$
M3 Ochiai	$\sqrt{x_1x_2} + y\sqrt{(1-x_1)(1-x_2)}$
M4 Jaccard	$\frac{x_1+x_2}{x_1+x_2-x_1x_2-y\sqrt{x_1(1-x_1)x_2(1-x_2)}} - 1$
M5 Sokal & Sneath 2	$\frac{2}{3} \cdot \frac{x_1+x_2}{2x_1+2x_2-3x_1x_2-3y\sqrt{x_1(1-x_1)x_2(1-x_2)}} - \frac{1}{3}$
M6 Czesanowski	$2 \cdot \frac{x_1x_2+y\sqrt{x_1(1-x_1)x_2(1-x_2)}}{x_1+x_2}$
M7 Psi	$y(0 \leq y \text{ の場合}), -y(y < 0 \text{ の場合})$
M8 Kulczynski	$\frac{1}{2} \cdot \frac{x_1+x_2}{x_1+x_2-2x_1x_2-2y\sqrt{x_1(1-x_1)x_2(1-x_2)}} - \frac{1}{2}$
M9 Yule's Q	$\frac{y}{2(1+y^2)\sqrt{x_1(1-x_1)x_2(1-x_2)}+(1-2x_1)(1-2x_2)y}$
M10 Russel & Rao	$x_1x_2 + y\sqrt{x_1(1-x_1)x_2(1-x_2)}$
M11 Rogers & Tanimoto	$\frac{2}{1+x_1+x_2-2x_1x_2-2y\sqrt{x_1(1-x_1)x_2(1-x_2)}} - 1$
M12 χ^2 metric (*2)	$\frac{1}{m_{k\bullet}}(\frac{1}{x_1} + \frac{1}{x_2} - 2 - 2y\sqrt{\frac{(1-x_1)(1-x_2)}{x_1x_2}})$
M13 Simple matching	$1 - x_1 - x_2 + 2x_1x_2 + 2y\sqrt{x_1(1-x_1)x_2(1-x_2)}$
M14 Sokal & Sneath 1	$2 - \frac{2}{2-x_1-x_2+2x_1x_2+2y\sqrt{x_1(1-x_1)x_2(1-x_2)}}$
M15 Hamann	$(1-2x_1)(1-2x_2) + 4y\sqrt{x_1(1-x_1)x_2(1-x_2)}$
M16 主効果の相乗平均	$\sqrt{x_1x_2}$

*1. x_1, x_2 はクロス集計表の行と列の主効果で、 y はクロス集計表の交互作用統計量。

*2. χ^2 metric については行和 $m_{k\bullet}$ が k に依存せず全て等しいと仮定した距離 (非類似係数) を記述。

量による影響より主効果の影響を強く受けて距離が長くなる。このため、タイプ III の類似係数は大規模疎行列をクラスタリングするために適さないと推測できる。

3.2.2 クラスタバリデーション

教師あり分類の妥当性は、真の分類との一致度で評価できる。しかし、本論では教師無し分類を扱うため、Milligan (1996) や Gordon (1999) の示すクラスタバリデーションの手順に従って表 3-5 の 5 つの基準を用いる。再現性・均等性・外的基準・内的基準の 4 評価は定

表 3-4. 主効果と交互作用統計量と類似係数の関係式 - その 2

名称	$x_1 = x_2 (= x)$ の場合の 類似係数の算式 $f(x, y)$ (*1)	$\left. \frac{\partial f(x, y)}{\partial y} \right _{x=0}$	タイプ
M1 交互作用統計量	y	1	I
M2 Cohen's Kappa	y	1	I
M3 Ochiai	$1 - (1 - x)(1 - y)$	1	II
M4 Jaccard	$\frac{2}{1 + (1 - x)(1 - y)} - 1$	$\frac{2}{(2 - y)^2}$	II
M5 Sokal & Sneath 2	$\frac{4}{3} \frac{1}{1 + 3(1 - x)(1 - y)} - \frac{1}{3}$	$\frac{4}{(4 - 3y)^2}$	II
M6 Czesanowski	$1 - (1 - x)(1 - y)$	1	II
M7 Psi	y ($0 \leq y$ の場合) $-y$ ($y < 0$ の場合)	1 ($0 < y$ の場合) -1 ($y < 0$ の場合)	I
M8 Kulczynski	$\frac{1}{2} \frac{1}{(1 - x)(1 - y)} - \frac{1}{2}$	$\frac{1}{2(1 - y)^2}$	II
M9 Yule's Q	$\frac{y}{2x(1 - x)(1 + y^2) + (1 - 2x)^2 y}$	0	III
M10 Russel & Rao	$x^2 + x(1 - x)y$	0	III
M11 Rogers & Tanimoto	$\frac{2}{1 + 2x - 2x^2 - 2x(1 - x)y} - 1$	0	III
M12 χ^2 metric (*2)	$\frac{2}{m_{k\bullet}} \frac{(1 - x)(1 - y)}{x}$	$-\infty$ に発散	III
M13 Simple matching	$1 - 2x + 2x^2 + 2x(1 - x)y$	0	III
M14 Sokal & Sneath 1	$2 - \frac{1}{1 - x + x^2 + x(1 - x)y}$	0	III
M15 Hamann	$(1 - 2x)^2 + 4x(1 - x)y$	0	III
M16 主効果の相乗平均	x	0	III

*1. x_1, x_2 はクロス集計表の行と列の主効果で、 y はクロス集計表の交互作用統計量。

*2. χ^2 metric については行和 $m_{k\bullet}$ が k に依存せず全て等しいと仮定した距離
(非類似係数) を記述。

量的な基準で、最後の解釈可能性は定性的な基準である。特に定量的な基準を用いる際には、各評価値の絶対水準の評価に加えて、類似係数間の相対的な評価順を確認する。

表 3-5. クラスタバリデーションの 5 つの基準

1. 再現性	分類を再現できる.
2. 均等性	各分類の大きさが均等.
3. 外的基準	一般に社会で受け入れられている分類に整合する.
4. 内的基準	消費者分類 × 製品分類の購買度数表に強い交互作用がある.
5. 解釈可能性	消費者と製品の分類、およびその対応を自然に解釈ができる.

再現性

階層的クラスター法では最初に作成されるグループが次に作成するグループの選択に影響を与える。このため、クラスタリングの初期にグループを形成する少数の観測対象の振る舞いに分類結果が影響を受ける懸念がある。そこで、階層的クラスター法について分類の再現性を評価する。評価手法は Milligan (1996) と Gordon (1999,p184) の手順に従って 2-フォールド交差確認法で評価する。

2-フォールド交差確認法の手順は以下の通り。

1. 観測対象を乱数を使って 2 等分し、それぞれ集合 A, B とする。
2. 集合 A を n 個のクラスター C_1, \dots, C_n に分類する。その際に C_i と集合 B の各点との距離も計算する。
3. 集合 B の各点を最も距離が近い C_i に分類する。
4. 集合 A と B を入れ替えて 2 と 3 の手順を繰り返し、もう一つの分類 C'_1, \dots, C'_n を得る。
6. 2 つの分類 $\{C_i\}$ と $\{C'_i\}$ の一致度をランドの修正指標で評価する。
5. 以上の手続きを N 回繰り返す。

ランドの修正指標はラベルに自明な一対一対応が無い分類の一致の評価に用いる指標で、ランドの指標 (Rand 1971) に Hubert & Arabie (1985) が修正を加えている。この指標について Milligan & Cooper (1986) は同じ目的で使う他の指標に比べて優れている事を示している。しかし、この指標値の確率密度関数の形状は、2 つの分類の対応表の周辺度数に影響を受けるため複雑で、有意水準を評価するのが難しい。そこで 50 回の交差確認法で得た 100 通りの分類からクラスターの大きさの分布の平均を求め、これを周辺度数として 2 つの独立な

分類を乱数で1万通り作成し、この対応表のランドの修正指標の分布を求める。そして、実際の分析で得た50回の指標値の平均から1標準偏差を引いた値の水準値が先のシミュレーションで得た分布のどこに位置するかを経験的パーセンタイル点を求めて、分類の独立性の検定のp値の代用とする。1標準偏差を引いて保守的に評価した理由は、2-フォールド交差確認法ではクラスター連結による新しい距離の更新のステップが半数となるため、空間の歪みの効果も半分となる事を考慮している。

均等性

実務的な観点から、分類の大きさの均等さを評価する。100個の観測対象を98個と1個と1個に3分類する事を分析者は望まないであろう。その結果が正しいか否かに関係無く、単に実務的な観点から均等な大きさの分類が望ましいと考える。均等さの指標は、観測対象を n 分類したグループを要素数の降順に並べ、上位半分のグループに含まれる要素数が全体に占める比率(%)で評価する。ただし、先の2-フォールド交差確認法で作成した100通りの分類について比率を計算し、その平均値を最終的な均等性の評価値とする。

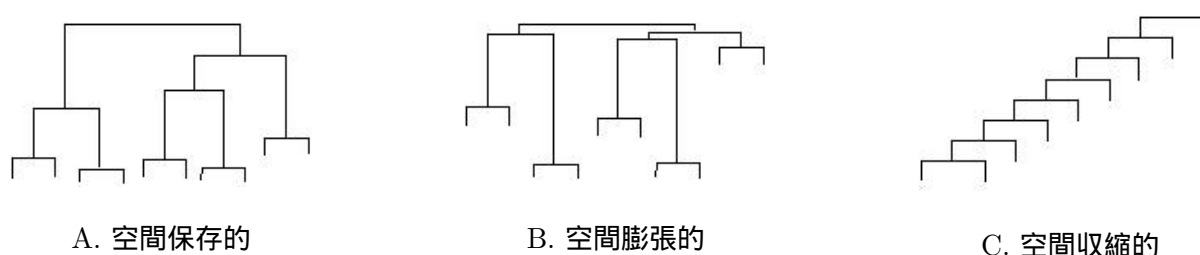


図 3-1. デンドログラムの類型

ところで、階層的クラスタリング法を用いた場合の均等性は、デンドログラムの形状と密接な関係がある。そして、類似係数の主効果のウェイトがデンドログラムの形状に影響を与えると予想する。そこで、均等性の評価ではデンドログラムの形状も併せて判断する。図3-1のデンドログラムは典型的な3類型で、それぞれA空間保存的、B空間膨張的、C空間収縮的と呼ばれる。これらの類型の要因には、クラスタリング法の選択 (Lance & Williams, 1967; DuBien & Warde, 1979) と、データの性質 (Yadohisa, Takeuchi, & Inada 1999) の2つがあると知られている。

類似係数の主効果のウェイトが大きくなると、類似係数が観測対象に固有の値となり、分類の階層構造も直線的なランキングに近くなる。このためデンドログラムは空間収縮的な

形状を示し、分類の大きさは均等にならないと予想する。デンドログラムの種類の判断は、2-フォールド交差確認法では無く通常のクラスタリングの結果に対して行う。また種類の判断は筆者が定性的に行う。

外的基準

観測対象の分類として広く受け入れられている分類を外的基準として、この分類に似ている程度を評価する。外的な分類は唯一の手本では無いので分析結果と一致する必要は無いし、分類数も異なるが、分析結果と無関係である事は望ましくない。そこで外的基準の分類と分析結果の分類との対応表を作成し、この対応表の独立性の検定指標であるカイ二乗統計量を外的基準の評価とする。

内的基準

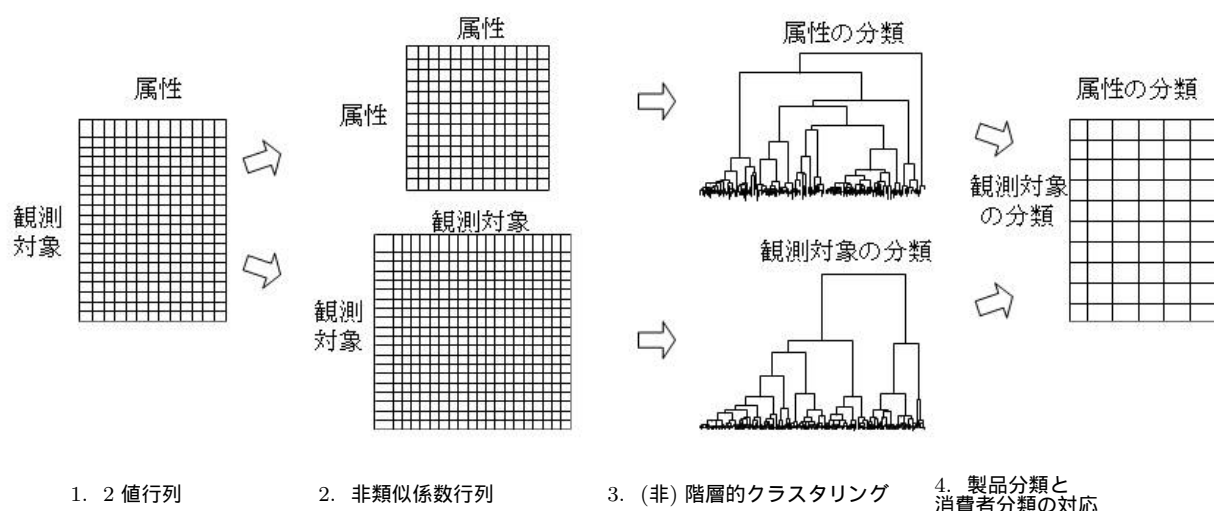


図 3-2. 分類の手順

内的基準は、意図した分類が作成できたかどうかを外的基準に頼らずに評価する基準である。類似係数を用いた分類を行う意図は、属性値のパターンが似ている観測対象をグループ化する事である。よって、観測対象のグループ数が属性の数に対して十分小さい時、観測対象のグループに対応した属性のグループも存在すると考えられる。そこで内的基準を、図 3-2 の通り、観測対象と属性の各グループの対応による製品カテゴリー分類と消費者セグメント分類との対応表の独立性の検定指標とする。属性の分類は、2 値行列 M の行と列の役

割を逆にして、観測対象を分類する場合と同様に作成する。そして観測対象と属性の分類を対応させて、製品と消費者分類の対応表を作成する。

解釈可能性

観測対象や属性に意味のあるラベルを与え、それらのラベルが同じ分類または違った分類に配置される事を、分析結果とは別の知見に照らして定性的に妥当な解釈が可能かを評価する。本実証分析の応用先である消費者行動分析や企業のマーケティング戦略評価においては、分析結果の利用者が市場構造についての仮説を持っている。したがって、記述的なクラスタリング手法により導出された分類軸やセグメントは、その仮説に沿って解釈できることが極めて重要な要件となる。そこで本論では観測対象について多くの知見を持つデータ提供者に、解釈可能性の評価を求める。

3.2.3 クラスタリング手法

類似係数と相性の良いクラスタリング手法の組合せがあると、クラスタリング手法の選択によって検証結果が影響を受ける懸念がある。そこで非階層的クラスタリング法を含む5種類のクラスタリング手法を検証に用いる。

表 3-6. クラスタリングの手法

	クラスタリングの手法	α_i	β	γ
非階層的	1 PAM 法			
階層的	2 ウォード法	$\frac{n_i+n_k}{n_+}$	$-\frac{n_k}{n_+}$	0
	3 群平均法	$\frac{n_i}{n_i+n_j}$	0	0
	4 最長距離法	$\frac{1}{2}$	0	$\frac{1}{2}$
	5 最短距離法	$\frac{1}{2}$	0	$-\frac{1}{2}$

表中の n_i は分類 C_i の要素の個数、また $n_+ \equiv n_i + n_j + n_k$ とする。

分類間の非類似距離 $d(C_i, C_j)$ が所与のとき、 C_i と C_j を併合した新しい分類 $C_i \cup C_j$ と他の任意の分類 C_k との非類似係数は、表中 α_i, β, γ を用いて次の通り。

$$d(C_i \cup C_j, C_k) \equiv \alpha_i d(C_i, C_k) + \alpha_j d(C_j, C_k) + \beta d(C_i, C_j) + \gamma |d(C_i, C_k) - d(C_j, C_k)|$$

PAM(Partitioning Around Medoids) 法は Kaufman & Rousseeuw (1990) によって提案された非階層的クラスタリング手法で、最適な分類を求めるアルゴリズムが k-means に似ているが、入力データが三角不等式を満たさない非類似係数でも使えるという利点がある。

ウォード法、群平均法、最長距離法、最短距離法の4つの手法は、最も近い2つの分類を逐次併合して新しい分類を作成する階層的なクラスタリング手法である。表 3-6 の通り、新しく併合した分類と他の分類との距離を決めるパラメーター α_i, β, γ に違いがある。Lance & Williams (1967) と DuBien & Warde (1979) は、表 3-6 のパラメーターの差異がクラスタリングの階層構造における空間の歪みの原因となり、図 3-1 の通りデンドログラムの3類型が生じる事を示している。

3.3 実証分析と結果

教師なし分類の妥当性を検証するクラスターバリデーションでは、多面的な基準による検証を必要とする。本研究では、クラスタリング法の選択に依存しない類似係数の選択基準に関心があるため、類似係数の選択とクラスタリング法の全ての組み合わせについて注意深く検証する。本文に記載しない結果は付表に記す。

3.3.1 データ

大手セル・レンタルCDチェーン店から、個人が特定できないように加工した上で提供頂いた音楽CD販売のID付POSデータを用いる。データは2002年11月1日から2003年12月21日までの約13ヶ月間の、日本の大都市にある5店の計707,085件の購買履歴となっている。前処理として個別タイトルをアーティストに名寄せし、20人以上の消費者に購入されたアーティストと10枚以上のCDを購入した消費者だけからなるデータセットに絞り込む。この結果、655アーティスト×3,577消費者の54,517件の購買履歴を分析対象とし、これ以降、アーティストで名寄せしたCDを便宜的に製品と呼称する。購買履歴と2値行列 $M = \{m_{ij}\}$ の関係は、製品 i を消費者 j が購入した履歴がPOSデータにある場合のみ $m_{ij} = 1$ 、それ以外は $m_{ij} = 0$ とする。したがって、消費者 j が同一製品を複数購入したり、繰返し購買した場合も $m_{ij} = 1$ とする。

表 3-7. 交互作用統計量と主効果との相関係数

非類似係数	製品		消費者	
	M1 交互作用統計量	M16 主効果の相乗平均	M1 交互作用統計量	M16 主効果の相乗平均
M1 交互作用統計量	1.00	0.01	1.00	0.09
M2 Cohen's Kappa	0.96	0.01	0.94	0.04
M3 Ochiai	0.90	0.38	0.94	0.15
M4 Jaccard	0.90	0.34	0.95	0.11
M5 Sokal & Sneath 2	0.90	0.32	0.96	0.10
M6 Czesanowski	0.90	0.35	0.94	0.11
M7 Psi	0.88	0.27	1.00	0.13
M8 Kulczynski	0.80	0.27	0.95	0.10
M9 Yule's Q	0.58	0.19	0.64	0.15
M10 Russel & Rao	0.50	0.70	0.88	0.41
M11 Rogers & Tanimoto	0.07	-0.90	0.12	-0.94
M12 χ^2 metric	0.07	0.76	0.34	0.45
M13 Simple matching	0.07	-0.90	0.12	-0.94
M14 Sokal & Sneath 1	0.07	-0.89	0.11	-0.94
M15 Hamann	-0.07	0.90	-0.12	0.94
M16 主効果の相乗平均	0.01	1.00	0.09	1.00

表中の各非類似係数の M1 から M15 の番号は、製品についての交互作用統計量との相関係数の降順に一致させている。その相関の水準は、1.0 に近い M1 ~ M8 と、0.5 付近の M9 ~ M10 と、0.0 に近い M11 ~ M16 の 3 グループに分かれる。

各類似係数と交互作用統計量、また主効果の相乗平均との相関係数を表 3-7 にまとめた。2 種類の類似係数行列 $S^1 = \{s_{i,j}^1\}$ と $S^2 = \{s_{i,j}^2\}$ の相関を計算する際は、それぞれ $s_{i,j}^k (k = 1, 2)$ を第 $i + (j - 1) \times m$ 要素に持つ列ベクトルを作成し、ベクトル間の相関係数として求める。

各類似係数の記載順の M1 から M16 は、製品間の類似係数について交互作用と相関の高い順にラベル付けしている。交互作用と主効果の相乗平均との相関係数は 0.01(製品間) と 0.09(消費者間) となり、2 つの類似係数はほぼ独立となっている。16 個の類似係数は、交互作用統計量と相関が高い M1 ~ M8 までのグループと、交互作用統計量とはほぼ無相関だが主効果の相乗平均と相関が高い M11 ~ M16 のグループ、およびどちらともいえない M9 と

M10 の 3 個のグループに分かれる。分類に適切な類似係数は交互作用統計量と相関が高いという予測を、分類結果の妥当性の評価順位が M1 から M16 のラベルの順となる事で検証する。

クラスターバリデーションに関する検証は、クラスタリング法の選択に結果が依存しないように、16 種類の類似係数と 5 種類のクラスタリング法の組合せで行う。また分類数については敢えて分類数を決める理由と基準に乏しいので、敢えて最適な分類数を一つ決める事はせず、観測対象と属性ともに 10,15,20,25,30 分類の 5 通りについて検証を行い、分類数の変化による分類結果の相違を確認する。

3.3.2 再現性の検証

階層的クラスタリング法について、50 回の 2-フォールド交差確認法により分類の再現性を評価する。分類の一致度はランドの修正指標で評価し、試行の平均値を分類数 (10,15,20,25,30 分類) 毎に求める。分析の結果、各分類数の評価とも概ね同じ傾向を示したので、表 3-8 に 15 分類の場合の指標の有意水準をまとめている。最長距離法と M1 交互作用統計量または M2 Cohens's Kappa の組み合わせを除き、最長距離法と最短距離法は再現性に乏しいと分かる。またワード法では M15 Hamann を除き再現性があるが、群平均法では M1 交互作用統計量との相関が低い M10 ~ M11 および M14 ~ M16 で再現性が乏しかった。

考察として、再現性が乏しくなるのは、分類の全体が少数の観測対象の振る舞いに左右されるからで、その要因はデンドログラムの形状に関係すると考える。図 3-1 の空間膨張的な場合は、併合するグループの決定がわずかな差で決まるため、分類結果が不安定になり易い。また空間収縮的な場合は、最初に形成されるグループが次に併合されるグループの選択に影響するため、最小のグループ形成が全体の分類に影響する。最長距離法と最短距離法の再現性が全般に乏しくなったのは、クラスタリング法の選択に要因であり、また群平均法において主効果と相関の高い類似係数の再現性が乏しくなるのは、類似係数の選択に要因があると考ええる。具体的なデンドログラムの類型は、次の均等性の検証で確認する。

3.3.3 均等性の検証

分類を構成する要素数の偏りを示す指標として、各分類を構成する要素数の多い上位半数の分類が全体の要素数に占める比率を計算する。例えば、100 個のメンバーを要素数が

表 3-8. 再現性の検証: ランドの修正指標と経験的パーセンタイル点

非類似係数	ワード法			群平均法			最長距離法			最短距離法	
	ランドの 修正指標	パーセン タイル点		ランドの 修正指標	パーセン タイル点		ランドの 修正指標	パーセン タイル点		ランドの 修正指標	パーセン タイル点
M1 交互作用統計量	0.35	0.0%	***	0.52	0.0%	***	0.15	0.0%	***	0.07	39%
M2 Cohen's Kappa	0.32	0.0%	***	0.49	0.0%	***	0.17	0.0%	***	0.04	39%
M3 Psi	0.36	0.0%	***	0.51	0.0%	***	-0.15	100%		0.08	40%
M4 Sokal & Sneath 2	0.31	0.0%	***	0.46	0.0%	***	-0.15	100%		0.06	40%
M5 Jaccard	0.31	0.0%	***	0.45	0.0%	***	-0.15	100%		0.05	43%
M6 Czesanowski	0.32	0.0%	***	0.45	0.0%	***	-0.15	100%		0.07	42%
M7 Kulczynski	0.32	0.0%	***	0.46	0.0%	***	-0.15	100%		0.07	44%
M8 Ochiai	0.33	0.0%	***	0.49	0.0%	***	-0.14	100%		0.05	41%
M9 Yule's Q	0.38	0.0%	***	0.40	0.0%	***	-0.12	100%		-0.01	100%
M10 Russel & Rao	0.39	0.0%	***	0.02	85%		-0.17	100%		-0.02	26%
M11 chi2 metric	0.39	0.0%	***	-0.02	27%		-0.07	100%		-0.02	26%
M12 Rogers & Tanimoto	0.62	0.0%	***	0.10	4%	*	0.25	4%	*	-0.02	100%
M13 Sokal & Sneath 1	0.58	0.0%	***	0.10	4%	*	0.21	19%		-0.01	100%
M14 Simple matching	0.59	0.0%	***	0.08	30%		0.22	12%		-0.02	100%
M15 Hamann	-0.02	26%		-0.02	26%		0.04	7%		-0.02	26%
M16 主効果の相乗平均	0.57	0.0%	***	-0.02	25%		-0.02	27%		-0.02	26%

優位水準 ***:0.1% **:1% *:5%

ランドの修正指標は最大値が 1 で、偶然一致と同じ水準で 0 の値をとる。

経験的パーセンタイル点は、モンテカルロシミュレーションによって得たランドの修正指標の分布から求める。

80,5,5,3,2,1,1,1,1,1 のクラスに 10 分類した場合には、 $(80 + 5 + 5 + 3 + 2)/100 = 95(\%)$ と計算する。100%に近い値は、約半分の分類が空に近い事を示している。また、完全に一樣ならば 50%となる。階層的クラスター法を用いる場合は、値の信頼性を高めるために 50 回の交差確認法で作成した 100 通りの分類の平均値を求める。

表 3-9 に 15 分類の検証結果をまとめた。また同じく表 3-9 に、デンドログラムの 3 類型への分類を筆者が判断して記載した。図 3-3 の実例の通り、ほとんど全てのデンドログラムが明確な特徴を示し、判断に迷う事はなかった。図 3-4 の箱ひげ図は、交差確認法による各分類に含まれる要素数の変動を、横軸に $i(= 1, \dots, 15)$ 番目の大きさの分類を置いて示したも

表 3-9. 均等性の検証: 分類の種類と要素数の多い上位半数のグループに含まれる要素数の
構成比

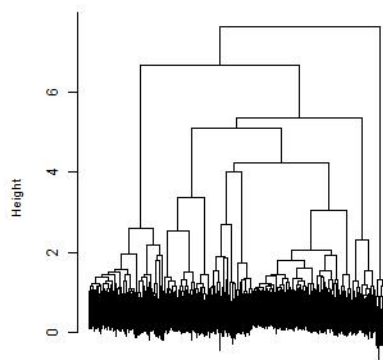
dissimilarity	PAM 法	ウォード法		群平均法		最長距離法		最短距離法	
	比率%	類型	比率%	類型	比率%	類型	比率%	類型	比率%
M1 交互作用統計量	66	A	75	B	85	B	64	C	99
M2 Cohen's Kappa	60	A	75	B	81	B	62	C	99
M3 Ochiai	69	A	73	B	86	B	82	C	99
M4 Jaccard	62	A	76	B	85	B	82	C	99
M5 Sokal & Sneath 2	62	A	78	B	85	B	83	C	99
M6 Czesanowski	63	A	74	B	85	B	82	C	99
M7 Psi	67	A	77	B	86	B	82	C	99
M8 Kulczynski	67	A	80	B	85	B	82	C	99
M9 Yule's Q	61	A	66	C	76	B	81	C	99
M10 Russel & Rao	77	A	83	B/C	89	B	84	C	99
M11 Rogers & Tanimoto	99	C	98	C	99	C	99	C	99
M12 χ^2 metric	99	C	96	C	99	C	95	C	99
M13 Simple matching	99	C	98	C	99	C	99	C	99
M14 Sokal & Sneath 1	99	C	98	C	99	C	99	C	99
M15 Hamann	99	C	99	C	99	C	93	C	99
M16 主効果の相乗平均	99	C	98	C	99	C	99	C	99

分類の種類 A,B,C は図 1 の通り A. 空間保存的,B. 空間膨張的,C. 空間収縮的を示す。

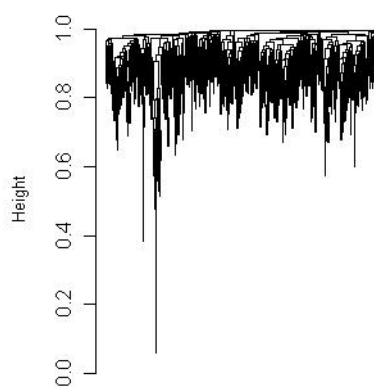
表中の比率 (%) は、製品を 15 グループに分類して要素数の降順にグループを並べ、最初の半分のグループに全体の要素数の何パーセントが含まれるか構成比を計算した値。ただし、第 8 グループについては要素数の半分を上位グループに含めた。

のである。このグラフから、要素数の分布はどのデンドログラムの種類でも比較的安定している、上記の均等性の指標も同様に安定していると考えられる。また表 3-9 に記載した以外の条件である、10,20,25,30 分類と消費者の分類についても評価したところ同様の傾向を示した。

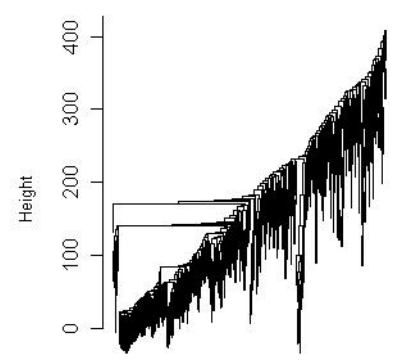
表 3-9 より、空間収縮的なデンドログラムと均等性の欠如は共通した現象であると分かる。また、主効果の相乗平均と相関が高い M11 ~ M16 類似係数は、全順序の性質を持つ類似係数であるために、空間収縮的なデンドログラムを発生させていると考える。ところで、非階



M1. 交互作用統計量
× ウォード法
空間保存的

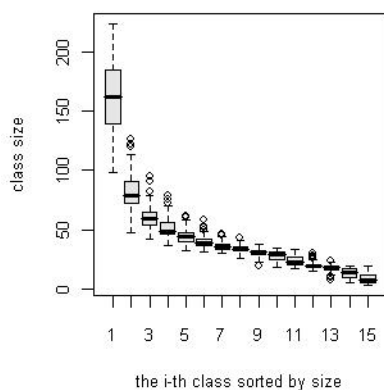


M4. Jaccard 係数
× 群平均法
空間膨張的

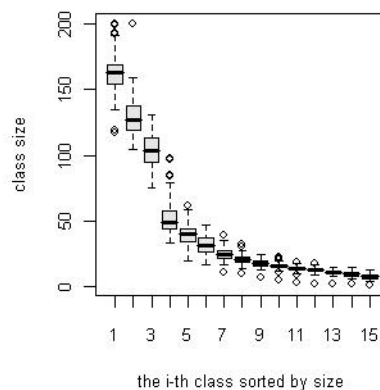


M12. χ^2 metrix
× 最長距離法
空間収縮的

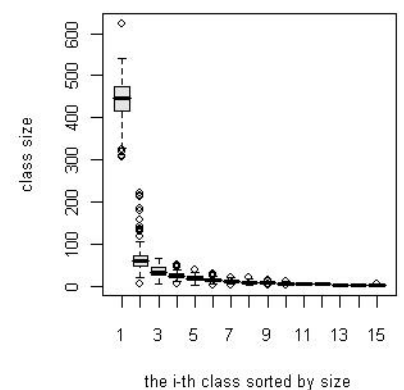
図 3-3. デンドログラムの空間の歪みの類型



M1. 交互作用統計量
× ウォード法
空間保存的



M4. Jaccard 係数
× 群平均法
空間膨張的



M12. χ^2 metrix
× 最長距離法
空間収縮的

2-フォールド交差確認法で 15 分類を 50 回作成し、各回において分類に含まれる要素数（縦軸）を降順に並べて（横軸）、値の分布を箱ひげ図に示している。

図 3-4. グループに含まれる要素数の分布の箱ひげ図（15 分類の場合）

層的手法の PAM 法は非階層的な手法より均等な分類を作成している。M1 ~ M10 の非類似係数について一つの例外を除き上位半分の占有率が階層的手法より小さく、PAM 法は均等な分類を作り易い事が分かる。M1 交互作用統計量と M2 Cohen's Kappa は他の類似係数に比べて最長距離法との組合せにおいて上位半分の構成比が 20%ほど低い。この 2 つの類似係数は、偶然一致を差し引くためプラスの交互作用と同様にマイナスの交互作用も適正に評

価できるので、最長距離法と相性が良いと考察する。

3.3.4 外的基準の検証

表 3-10. 外的基準の分類と分析の結果得られた分類との一致数の例

ジャンル	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B13	B14	B15	計
1 邦 ポップ	19	104	35	22	40	15	36	42	38	0	0	0	0	0	0	351
2 ロック & ポップ	0	0	0	0	0	0	0	66	2	26	28	1	0	19	34	176
3 クラブ / ダンス	0	0	0	0	0	0	0	3	13	0	0	0	0	0	0	16
4 ハードロック / メタル	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	6
5 邦 ジャズ	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
6 洋 ジャズ	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2
7 キッズ	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
8 アニメ / ゲーム	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	3
9 ブラック / ソウル	0	0	0	0	1	0	0	8	1	0	2	33	0	0	0	45
10 ワールド	0	0	0	0	0	0	0	1	0	0	2	0	0	0	0	3
11 ヒップホップ / ラップ	0	0	0	0	0	0	0	3	0	0	0	4	32	0	0	39
12 イーギリスニング	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2
13 クラシック Classic	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	3
14 レゲエ	0	0	0	0	2	0	0	3	0	0	0	1	1	0	0	7
計	19	107	35	22	43	15	36	141	54	26	32	39	33	19	34	655

行方向: 外的基準は POS データと共にデータ提供者から提示された製品属性。

列方法: B1 ~ B15 のラベルは交互作用統計量とワード法の組合せで作成した 15 分類。

P O S データの提供者から受領した音楽 C D の流通や管理目的に使用する音楽ジャンル区分を外的な基準として、購買履歴からクラスタリング法で作成した分類と、外的基準との一致の程度を検証する。対応の例として、表 3-10 に外的な音楽ジャンル区分を行方向に、交互作用統計量とワード法の組み合わせで作成した分類を列方向として、各分類に該当する製品数を表に示している。

外的基準の音楽ジャンル区分の特徴として、表 3-10 の 1 行目と 2 行目に相当する「1 邦ポップ」と「2 ロック & ポップ」に該当する製品が多く、分類対象 655 製品の約 8 割の 527 製品がこの 2 ジャンルに区分されている。音楽ジャンルが店舗内の棚割りに応用される場合、製品数の多い分類は再分類されて近い配置になっている方が目的の製品を探す消費者にとって都合が良いが、外的基準にはそのような再分類が無い。これに対して、交互作用統計

量とワード法の組み合わせで作成した分類は、外的基準で製品数の多い「1 邦ポップ」と「2 ロック & ポップ」を「女性アーティスト」や「ビジュアル系」などに再分類している（表 3-11、Appendix 表 A-1-5 アーティスト名一覧、参照）。「女性アーティスト」は、浜崎あゆみや Every Little Things のように女性ボーカルの魅力が強いアーティスト分類であり、「ビジュアル系」は globe や L'Arc ~ en ~ Ciel のようにグループとしての魅力が強いアーティスト分類として、より均等で解釈可能な分類となっている。同時に、この 2 分類以外は多数の分類に細分する事なく、14 行目の「14 レゲエ」を除き 1 分類か 2 分類への対応を与えている。このため 2 種類の分類に強い関係があると確認できる。このように交互作用統計量とワード法の組み合わせで作成した分類は、外的基準と整合的で、結果として表 3-10 には 0 の値が多くなっている。

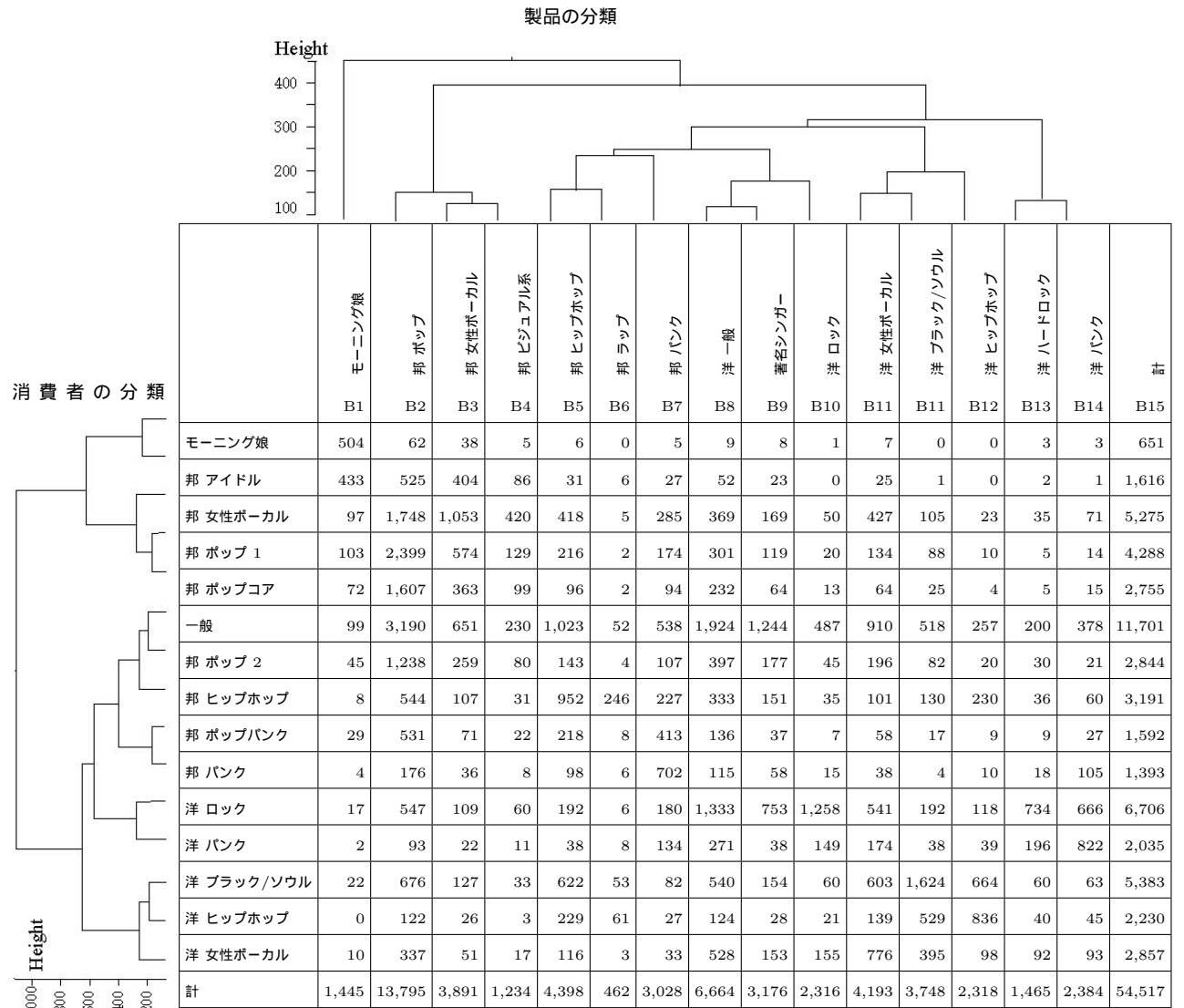
表 3-10 の行方向と列方向の分類の独立性をカイ二乗統計量で検証するのは、セルに 0 の値が多いという問題があるが、参考として 16 種類の非類似係数と 5 種類のクラスタリング法の組合せについて対応表のカイ二乗統計量を求めて相対的な比較を行ったところ、M1 ~ M10 の類似係数を選択した際のカイ二乗統計量は、M11 から M16 の類似係数を選択した場合に比べて総じて高く良い値を示した。またクラスター法の比較については、PAM 法とワード法と群平均法が他の 2 つの方法に比べてカイ二乗統計量が大きく良い結果を示した。

3.3.5 内的基準の検証

表 3-11 に、交互作用の非類似係数とワード法の組合せで作成した製品分類と消費者分類の対応表の一例を示す。表中の各セルの値は、該当する分類の 2 値行列 $M = \{m_{ij}\}$ の小計に相当している。また分類名は、分類メンバーのアーティスト名を基に筆者が命名している。

これらの製品分類と消費者分類の対応表のカイ二乗統計量を、外的基準の場合と同様に分析し、結果を表 3-12 にまとめた。消費者と製品それぞれ 5 通りの計 25 通りの分類数について計算し、表にはその集計値を記している。交互作用統計量と相関が高い類似係数に PAM 法かワード法を用いた組合せでカイ二乗統計量が 200 後半の高い値となり、好ましい結果を示している。特に M1 交互作用統計量、M8 Ochiai、M3 Psi、M2 Cohen's Kappa が上位を占め、カイ二乗統計量の順位の標準偏差も小さいため、順序関係も安定している事が分かる。

表 3-11. 製品分類と消費者分類の対応表の例



表中の値は購買履歴の該当数。ただし、同じ消費者による同一製品の繰返し購買は重複カウントしない。製品と消費者の分類名は筆者による命名。

3.3.6 解釈可能性の検証

ラベルが M1 ~ M8 の類似係数に PAM 法かワード法を適用した分類は、表 3-11 の製品分類と消費者分類の対応表と概ね同じ結果になった。それ以外の類似係数とクラスタリング法の組み合わせを用いた分類は、著しく均等性が欠如していたり、あるいは同じ音楽性を持つアーティストが別の分類に属する不自然さがあり、第三者の意見を伺うまでも無く解釈可能でない分類と評価した。

表 3-12. 内的基準の検証: 製品分類と消費者分類のカイ二乗統計量のランキング

非類似係数	$\chi^2/d.f.$ 統計量の順位の平均 (標準偏差)					$\chi^2/df.$ の平均				
	PAM 法	ウォード 法	群平均 法	最長 距離法	最短 距離法	PAM 法	ウォード 法	群平均 法	最長 距離法	最短 距離法
M1 交互作用統計量	1.3 (0.5)	1.8 (1.0)	4.6 (1.9)	2.0 (0.0)	5.0 (1.8)	266	294	149	113	3
M2 Cohen's Kappa	4.1 (1.0)	3.2 (1.2)	1.9 (0.3)	1.0 (0.0)	10.0 (0.0)	249	294	163	139	2
M3 Psi	3.1 (1.9)	3.2 (0.8)	9.0 (0.0)	10.9 (1.4)	4.8 (1.6)	252	293	139	22	3
M4 Sokal and Sneath 2	5.1 (1.2)	6.0 (0.8)	6.0 (1.2)	7.2 (1.0)	7.5 (1.4)	246	282	145	30	2
M5 Jaccard	6.0 (1.2)	5.9 (0.7)	7.0 (0.9)	5.8 (1.0)	7.2 (1.2)	243	283	145	31	2
M6 Czesanowski	6.1 (1.4)	6.2 (1.1)	5.3 (1.3)	6.0 (1.3)	7.1 (1.8)	239	284	146	31	2
M7 Kulczynski	8.0 (0.7)	7.9 (0.3)	6.4 (1.4)	6.8 (1.5)	6.5 (1.8)	231	278	146	30	2
M8 Ochiai	2.6 (1.2)	1.8 (0.5)	3.8 (1.0)	4.4 (0.7)	3.8 (1.1)	258	296	149	31	3
M9 Yule's Q	9.6 (0.5)	9.6 (0.5)	1.1 (0.3)	3.0 (0.0)	11.4 (0.5)	187	257	168	34	1
M10 Russel and Rao	9.0 (1.1)	9.4 (0.5)	10.0 (0.0)	8.8 (0.7)	2.2 (0.8)	221	252	34	28	3
M11 カイ二乗距離	13.7 (0.8)	11.0 (0.0)	11.0 (0.0)	10.7 (0.7)	1.0 (0.0)	80	183	9	22	8
M12 Rogers and Tanimoto	11.0 (0.0)	12.0 (0.0)	13.9 (1.1)	12.9 (0.7)	14.1 (0.9)	139	118	1	16	1
M13 Sokal and Sneath 1	12.7 (0.4)	14.0 (0.0)	13.6 (1.1)	12.4 (1.0)	13.7 (1.1)	117	106	1	16	1
M14 Simple matching	12.7 (0.4)	13.0 (0.0)	13.4 (1.0)	13.0 (1.2)	13.7 (0.8)	117	113	1	16	1
M15 Hamann	15.0 (0.0)	16.0 (0.0)	15.0 (1.7)	15.0 (0.0)	15.8 (0.6)	3	1	1	2	0
M16 主効果の相乗平均	16.0 (0.0)	15.0 (0.0)	14.0 (1.6)	16.0 (0.0)	12.2 (1.6)	0	3	1	1	1

表の値は、消費者と製品の分類数をそれぞれ 10,15,20,25 および 30 個の 5 通りで作成した計 5×25 パターンの対応表の $\chi^2/d.f.$ 統計量の降順の順位と、その順位の標準偏差である。ただし、各分類パターンの $\chi^2/d.f.$ は 2-フォールド交差確認法による 50 回のクラスタリングの平均値を採用している。

そこで表 3-11 のアーティストのクラスタリング構造についての解釈可能性のコメントを、職業的専門家であるデータ提供者に求めたところ、音楽 CD の店舗内の棚割に用いているジャンル区分と似ていて、邦楽と洋楽に分かれ、さらに細分している構造は違和感の無い自然な分類であるとの回答を得た。音楽のジャンルには盛衰があり、クラシックやジャズやロックの新譜が多かった頃に比べると、今日はポップの新譜が多くなっている。そうなる、消費者は多数のポップの新譜の全てに関心を払うことができなくなり、またアーティストの側は自身の付加価値を高めるために他のアーティストから差別化する利点が大きくなる。そこで、文学や絵画の流派がそうであるように、主流となった音楽ジャンルは再分化し易くなり、音楽ジャンルを改定する必要が生じる。実際の購買履歴から分類を作成する本研

究の手法は、市場に変化に即した解釈可能な分類を導き易いと考えられる。

3.3.7 結果

5つのクラスターバリデーションの検証分析から16種類の類似係数の適切さを相対的に評価すると、まず主効果の相乗平均と高い相関を持つM11～M16の指標グループは再現性と均等性で劣後し、解釈に適したデンドログラムの形状を示さない事が分かる。また外的基準と内的基準で、交互作用統計量と相関が高いM1～M8の類似係数のグループの下位に評価される。次にM9とM10の2指標は、再現性や均等性あるいは外的基準による評価で良好な結果もあるが、内的基準による評価がM1～M9の指標の下位に位置している。整理すると指標グループとして評価の良い順にM1～M8、M9とM10、M11～M16となる。

M1～M8の指標は概ね同じ有効性を示しているが、特にM1 交互作用統計量とM3 Ochiaiが外的基準と内的基準の評価で安定して上位にある。またM1 交互作用統計量とM2 Cohen's Kappaは最長距離法を用いた場合にも高い有効性を示し、交互作用統計量がマイナスの場合も適切に評価していると推測する。この結果と、16指標の有効性が全体として交互作用と相関の高い順に並んでいることを勘案すると、相対的に優れた指標はM1 交互作用統計量と結論できる。

M9以下の類似係数がクラスタリングに適さなかった事は、類似係数を主効果と交互作用統計量の関数として分類した表3-4の結果と整合する。大規模疎行列では主効果が小さくなり、このとき表3-4でタイプIIIと分類したM9以下の類似係数は交互作用統計量と無相関に近づく関数上の特徴を持っている。このため、これらの類似係数では、図3-3のカイ二乗距離の空間収縮的なデンドログラムのように、主効果の大小の順にクラスターの凝縮を導く。

3.4 考察と課題

3.4.1 考察

大規模疎行列データのクラスタリングに適した類似係数を実験的に評価し、また類似係数を交互作用統計量の導関数として評価した結果、クラスタリングに適した類似係数は主効果がゼロに近づいても交互作用統計量との相関を失わない類似係数である事を確認できた。

そのような特徴を持つ類似係数は、主効果の影響を受けない類似係数と、主効果と交互作用の影響を等しく受ける類似係数の2つのタイプに分類された。

類似係数を選択するためのクラスターバリデーションの評価基準として、再現性・均等性・外的基準・内的基準・解釈可能性について検証を行ったが、マーケティングの観点からは、データ提供者が用いる流通のための音楽ジャンル区分との対応である外的基準と、解釈可能性がとくに重要であると考えた。交互作用統計量を用いた分類と、外的基準との対応表は表3-10. に示した通り、外的基準による分類の過半数を占める「1 邦ポップ」と「2 ロック&ポップ」に対しては細分類を与え、それ以外については、ほぼ外的基準による分類の過半が本分析で作成した分類に対して1対1対応を示した。また、解釈可能性については、データ提供者から、邦楽と洋楽の2分類があり、それぞれがロック、ポップといった類似した構造に分類されており、違和感の無い自然な分類であるとの回答を得た。以上を踏まえて、第3章で作成した交互作用統計量を用いた分類は、識者の内面にある製品カテゴリー分類に近い結果になったと解釈できる。

階層的なクラスター分析において、空間収縮的か空間膨張的なデンドログラムは実用的でないと評価したが、製品の特徴が一つの評価軸で評価できるような性質を持つ市場であれば、製品を順序付けによって分類できるため、空間収縮的であろうと考えられる。例えば価格と品質が直線関係にあり、他の属性評価軸の無い市場は空間収縮的となろう。また、空間膨張的については、カテゴリーの凝集が生じ難く、製品カテゴリーを見出しづらい混沌とした市場と解釈できる。音楽CD市場は、これらの特徴にあてはまる市場では無く、交互作用統計量を用いて空間保存的な分類を得た事は、妥当な結果であったと解釈できる。

本研究の応用分野として、大規模疎行列データを扱う事が多いゲノム解析やマーケティングのPOSデータ解析があろう。記述的なデータ解析の第一段階として、交互作用統計量との相関を失わない類似係数を用いて観測対象の類似を評価する事は有益と考える。

3.4.2 課題

類似係数の算式の選択についての今後の課題としては、序論で言及したようにピアソンのカイ二乗統計量以外の交互作用を検討する事があろう。また、教師なし分類の妥当性を評価するクラスターバリデーションについては、まだ標準的な手続きや評価基準が定まっているとは言い難く、さらに議論を進めて評価の客観性を高める余地があると考えた。評価基準の

外的基準と解釈可能性に関して、消費者の側の評価を得ていない事も課題と言える。製品提供者と消費者の側では、異なる製品カテゴリーの類似関係を捉えているとも考えられるので、第3章の目的に照らすと、消費者側に評価を求める実験の意義があると思われる。

第4章 交互作用統計量に基づく製品選択の研究

4.1 はじめに

第3章では、クラスターバリデーションの評価基準を用いた検証により、行動面から製品カテゴリーの類似関係を推定する際には、交互作用統計量を用いたクラスター分析の分類結果が、他の類似係数よりも優位である事が分かった。第4章では、逆の経路として、交互作用統計量を用いた製品カテゴリーの類似評価を用いて、消費者の製品選択を説明するモデルを構成し、その妥当性を検証する。評価の手順として、実データと模擬データを用い、推定に用いた行動データより将来の購買が有意に予測できる事を検証する。提案モデルの構築を通して、製品選択における消費者の異質性の所在の設定を明示し、製品カテゴリーの類似関係が売上に対して及ぼす影響を評価する。

消費者が製品や提示された情報を処理する方法はさまざまであるが、その代表的なものに、ピースミール処理とカテゴリーベース処理がある（たとえば、Sujan,1985; Meyers-levy & Tybout, 1989）。前者は製品や情報を、それを構成する属性に分解して処理する方法であり、代表的なモデルに多属性態度モデルがある。それに対して、カテゴリーベース処理とは、提示された製品や情報を事前に消費者が知識としてもつ何らかの基準で作られたカテゴリーと照らし合わせることによって、対象を理解しようとする方式である。たとえば、市場にある新製品が投入されたときに、その新製品が、消費者の知識として形成されているあるカテゴリーとうまく合致するものであれば、その新製品は既存のカテゴリーのバリエーションとして理解されるというものである。カテゴリー化の方法としては、カテゴリーとカテゴリーメンバー間の帰属関係による方法、典型的やプロトタイプ、あるいは、ある目的の達成のために消費者が独自に創造するなどさまざまである（新倉 2001）。そして、今日のように、オンラインショップで購買可能な製品が多数あり、膨大な製品情報が提供されている状況下においては、ピースミール処理よりもカテゴリーベース処理を行うことが多いと考えら

れている（清水 1999）。このため、購買履歴から消費者の情報処理を分析する手法は、消費者がピースミール処理を行わない場合にも適用できる柔軟な手法となり得る。そこで本研究では、ピースミール処理を前提としない評価手法を提示する。

第4章では、消費者が同じプロダクトマップを持つという仮定に基づいて、製品の類似関係で消費者の購買を説明する数理モデルを構築する。そして、音楽CDの販売POSデータを用いて、そのモデルの購買予測力が有意である事を検証する。さらに、検証結果がデータセットや音楽業界固有の特徴に依存する限界を緩和するため、模擬のPOSデータを用いた検証も同様に行う。また、音楽業界のレコードレーベルを取り上げて、消費者の製品の類似関係の認識を利用したプロモーションについて議論する。具体的には、売上がプラスの交互作用を持つ2人のアーティストが期間併売され易い事を利用し、これらのアーティストを同じレコードレーベルという組織でプロモーションする利点を、モデルを用いて評価する。製品属性が消費者の製品選択に影響を与える場合に、製品属性がマーケティング変数となるように、類似関係が消費者の購買に影響を与える場合は、類似関係がマーケティング変数になり得ると考えられる。

4.2 先行研究

製品の類似関係と購買とを関連付ける先行研究として、「この商品を買った人はこんな商品も買っています」といった、レコメンドシステムに関する研究がある。オンラインショップは製品を陳列する物理的制約が無いため、多数の製品と情報を提供できる利点がある一方で、消費者の情報処理コストを増加させる懸念もある。そこで消費者毎に購入候補の考慮集合を提案し、購入候補の比較検討を支援するレコメンドシステム（Häubl & Trifts, 2000）が、有効なマーケティングツールとなる。大半のレコメンドシステムは、消費者が製品の類似関係を利用しており、このアプローチは、活性化拡散モデル（Collins & Loftus, 1975）やネットワークモデル（Aaker, 1996; Aaker, 2004）において消費者の関心がネットワーク上を拡散し、関連して購買が起きるという考えに沿っている。すなわち、ある製品を購入したり関心を持った消費者に、その製品と類似関係の高い別の製品をレコメンドして認知させる事で、次の購買の考慮集合（Shocker et al., 1991）にレコメンドする製品が加わる効果が期待できる。

レコメンドシステムのアルゴリズムのサーベイ研究である Adomavicius & Tuzhilin (2005)

表 4-1. レコメンドシステムのアルゴリズムの分類と手法

	ヒューリスティックベース法	モデルベース法
コンテンツベース フィルタリング	頻度に基づく評価、クラスタリング	潜在クラスモデル、クラスタリング、決定木、ニューラルネットワーク
協調 フィルタリング	近傍法（類似した利用者・アイテムの評価を利用）、クラスタリング、グラフ理論	ベイジアンネットワーク、クラスタリング、ニューラルネットワーク、線形回帰、確率モデル
ハイブリッド フィルタリング	上記手法を併用あるいは経験的に線形結合	上記手法をモデルとして結合

Adomavicius & Tuzhilin (2005) の表 2 を参考に作成。

と神嶋 (2007, 2008(1), 2008(2)) と芳賀 (2008) は、利用するデータの違いから、レコメンドシステムを表 4-1 の行の通り 3 分類している。まずコンテンツベースフィルタリングは、利用者（顧客）とアイテム（評価対象や商品）の属性を対応づける手法である。次の協調フィルタリングは、利用者のアイテムに対する評価履歴や購買履歴の類似関係を手掛かりとする手法である。最初に協調フィルタリングの理論的アイデアを提示したのは、米国のゼロックス研究所の Tapestry (Goldberg et al., 1992) とされる。また GroupLens (Resnick, Iacovou, Suchak, Bergstrom & Riedl, 1994) は実際に運用されたシステムとして知られている。3 番目の分類である属性の対応と類似関係の両者を併用する手法は、ハイブリッドフィルタリングと呼ばれる。Bruyn, Liechty, Huizingh & Lilien(2008) は、コンジョイント分析を利用したレコメンドシステムを提案している。

レコメンドシステムのもう 1 つの分類軸として、表 4-1 の列に対応するヒューリスティックベース法（あるいはメモリーベース法）とモデルベース法の 2 分類がある。両者の違いは、いわば設計プロセスの違いで、そのアルゴリズムやパラメータ推定に至る過程が経験的か、あるいは理論的なモデルに基づくかで区別される。本論で提案するモデルは、経験的でなく、また属性を用いていないという点で協調フィルタリングのモデルベース法に分類できる。

表 4-1 以外のレコメンドシステムの分類として、評価値が 5 得点評価のような順序データ

か、あるいは購買の有無のような 2 値データかという分類も重要となる。順序データの場合は、相関係数で類似関係を評価できるが、2 値データの場合は同様の指標が使えない。このため、全ての利用者の 0 / 1 データを 2 人の利用者（あるいはアイテム）についてクロス集計して求める類似係数が利用される。これまで多数の類似係数が提案されているが（Wedel & Kamakura, 2000; Albatineh et al., 2006）、類似係数の選定基準についての定説は無い。石田・西尾・椿（2011）は、類似係数を交互作用統計量と主効果の関数として記述し、類似係数の有効性が交互作用統計量との相関関係で解釈できる事を実証と合わせて示している。

類似係数の利用について、Iacobucci, Arabie & Bodapati (2000) は製品の類似評価に対してクラスター分析を行い、同様に消費者の類似評価に対してクラスター分析を行い、これを対応させてレコメンドシステムに応用する概念を示している。Ariely, Lynch & Aparicio (2004) は類似評価の高い消費者に限定して集計することで計算量を減らしてシステムの学習効率を改善できる事を示している。また、類似係数行列をヒルベルト空間の内積であるカーネルとみなし、サポートベクターマシンを用いて購買の有無を判別する推定法もある（Shawe-Taylor & Cristianini, 2004）。本論で提案するモデルを含め、これらの 3 つのアプローチの導出過程に関連は無いが、共に類似係数を用いて購買を説明するという共通点を持っている。

本論では、消費者 i が製品 k を過去に購入したか否かが、その後に消費者 i が他の製品 j を購入するか否かに影響を与えると考え、製品 j の購買確率の算式を条件付き確率として記述する。これは、Breese, Heckerman & Kadie (1998) が利用者 a のアイテム j への評価得点 $v_{a,j} (\in 0, \dots, m)$ の期待値を、

$$E(v_{a,j}) = \sum_{i=0}^m Pr(v_{a,j} = i | v_{a,k}, k \in I_a) \cdot i$$

ただし I_a は利用者 a の評価済みアイテムの集合、と条件付き確率で記述しているのと共通した設定である。 $Pr(v_{a,j} | v_{a,k})$ の計算については、利用者を潜在クラスモデルで分類して利用者数別の推定をグループ数別に減らす手法と、ベイジアンネットワークで回帰する 2 つの手法を比較検証している。潜在クラスモデルが製品と市場の構造を分析する競争市場構造分析（井上, 2001）でも利用されるように、モデルベース法と競争市場構造分析のアプローチには共通点が多い。これに対して本論では、類似係数として交互作用統計量を用いるため、条件付き確率を自然に記述できている。

4.3 モデル

4.3.1 モデルの意図

ある製品を購入するという経験を通して消費者の購買行動が変化し、次に購入する製品の購買確率が購入前に比べて増減する程度を、購買履歴を条件とする条件付き確率で記述する。モデルの設定として、消費者がA製品を買うという事象は確率的で、他のB製品を買うという可能性もあるが、もしA製品を購入すれば、その購買結果が次の購買確率の条件となり、B製品を買った場合とは異なる購買行動を示すと考える。このため、購買履歴の無い初期状態で全ての消費者が完全に同質であると設定しても、確率的に起きる購買経験を通して消費者が異質性を示すモデルを考える。

モデル化する変数として、個々の消費者と製品は区別するが、時間と購買数量については考慮せず、一定期間の購買の有無のみを区別する。このためモデルを適用可能な製品は、1回に1個ずつ購入する性質の製品とする。特に耐久財や本論で扱う音楽CDのように、繰り返し購買が起きない製品にモデルを適用する場合は、繰り返し購買となる推定確率をゼロで上書きして対応する。

モデルの制約として、消費者行動の知見をモデル化するための制約と、計算を簡易にするための2つの制約を設ける。第1の制約として、製品間の類似関係は全ての消費者に共通であるとする。これは、どの消費者にも市場における製品の位置関係が同じに見える事を意味している。例えば、主成分分析やコレスポンデンス分析を利用したプロダクトマップの作成において、消費者毎に異なるプロダクトマップを作成しないのと同じである。この制約は、数値計算において購買履歴の持つ情報量の少なさを補い、消費者×製品の全ての組み合わせの購買確率の推定を可能とする重要な役割を果たす。具体的には、各製品と消費者の販売数で基準化した2製品の購買の同時確率についての交互作用統計量が、全消費者で共通とする。交互作用統計量以外にも様々な類似係数が同時確率の指標となり得るが、様々な類似係数よりも交互作用統計量が消費者や製品のクラスタリング指標と優位である事を示した石田・西尾・椿(2011)の結果を踏まえ、交互作用統計量を用いた制約を置く。

もう一つの制約として、2製品の期間併売を複数製品の期間併売に拡張するための仮定を置く。すなわち、1個の製品の購買の有無を条件とする購買確率を、条件となる製品を取り替えて全製品について単純平均する計算手法を採る(図4-1参照)。このように複数の推定

モデルを結合する手法はコミッティ (committe) と呼ばれ (Bishop, 2006) ブースティングなどの良く知られた重み付けの最適化手法がある。また、カーネル法にサポートベクターマシン適用する手法も同様に、重み付けを最適化する手法である。提案モデルでは周辺確率が重みの役割を果たすため、重み付けの重複を避けるために簡易に単純平均を用いる。

提案モデルの利点として、購買確率を集計して売上を予想できる事がある。確率モデルとしないレコメンドシステムでは、各顧客に推奨する製品の順序を示す事はできても、売上増加効果といった経営に必要な評価ができないという課題があった。

4.3.2 理論モデル

モデルの記述のために消費者と製品の集合、および所与の期間の購買履歴を次の通り記す。

$$I = \{1, \dots, i, \dots, n\} \quad n \text{ 人の消費者の集合}$$

$$A = \{a_1, \dots, a_j, \dots, a_m\} \quad m \text{ 個の製品の集合}$$

$$U = \{u_{i,j} \mid \text{消費者 } i(\in I) \text{ が製品 } a_j(\in A) \text{ を 1 個以上購入した時に限り}$$

$$u_{i,j} = 1 \text{ さもなければ } u_{i,j} = 0\}$$

また、消費者 $i(\in I)$ が製品 $a, b(\in A)$ を 1 個以上購買する条件無しの周辺確率、条件付き確率、同時確率をそれぞれ次の通り記す。以降、1 個以上の購買数量の多少は区別しない。

表 4-2. 製品 a と b の購買に関する発生確率

		製品 b		
		購買する	購買しない	計
製品 a	購買する	$P_i(a \cap b)$	$P_i(a) - P_i(a \cap b)$	$P_i(a)$
	購買しない	$P_i(b) - P_i(a \cap b)$	$1 + P_i(a \cap b) - P_i(a) - P_i(b)$	$1 - P_i(a)$
	計	$P_i(b)$	$1 - P_i(b)$	1

$P_i(a)$ 消費者 i が製品 a を購買する周辺確率

$P_i(b|a)$ 消費者 i が製品 a を購買した条件下で、製品 b を購買する条件付き確率

$P_i(a \cap b)$ 消費者 i が製品 a と b を共に購買する同時確率

表 4-2 に、消費者 i の製品 a と b の購買に関する 4 通りの事象の発生確率をまとめている。このとき、度数の場合の交互作用統計量と同様に、表 4-2 における交互作用統計量を以下の通り定義する。表 4-2 の各確率に従い 2 製品の購買を各 r 回試行した場合の各事象の期待値は、表 4-2 の各セルの値に r を乗じた値となる。2 製品の購買の独立性を検定するピアソンのカイ二乗値 χ^2 は、

$$\chi^2 = (r-1)s_i^2(a, b)$$

$$\text{ただし、} s_i(a, b) = \frac{P_i(a \cap b) - P_i(a)P_i(b)}{\sqrt{P_i(a)(1-P_i(a))P_i(b)(1-P_i(b))}} \quad (4.1)$$

カイ二乗値 χ^2 を試行回数 r で基準化して交互作用の正負に配慮して平方根をとった (4.1) 式の $s_i(a, b)$ を、本論文では交互作用統計量と呼ぶ。このとき、消費者 i が製品 a を購買した場合に製品 b を購買する条件付き確率 $P_i(b|a=1)$ は、(4.1) 式を用いて

$$P_i(b|a=1) = \frac{P_i(a \cap b)}{P_i(a)}$$

$$= P_i(b) + s_i(a, b) \sqrt{\frac{1-P_i(a)}{P_i(a)} P_i(b)(1-P_i(b))} \quad (4.2)$$

同様に製品 a を購買しなかった場合に製品 b を購買する条件付き確率は

$$P_i(b|a=0) = \frac{P_i(b) - P_i(a \cap b)}{1-P_i(a)}$$

$$= P_i(b) - s_i(a, b) \sqrt{\frac{P_i(a)}{1-P_i(a)} P_i(b)(1-P_i(b))} \quad (4.3)$$

(4.2) 式と (4.3) 式をまとめて、

$$P_i(b|a) = P_i(b) \left\{ 1 + t_{i,a} s_i(a, b) \sqrt{\left(\frac{1-P_i(a)}{P_i(a)} \right)^{t_{i,a}} \frac{1-P_i(b)}{P_i(b)}} \right\} \quad (4.4)$$

$$\text{ただし、} t_{i,a} = \begin{cases} 1 & (a \text{ を購買する場合}) \\ -1 & (a \text{ を購買しない場合}) \end{cases}$$

(4.4) 式を変形すると

$$\frac{P_i(b|a)}{P_i(b)} - 1 = t_{i,a} s_i(a, b) \sqrt{\left(\frac{1 - P_i(a)}{P_i(a)}\right)^{t_{i,a}} \frac{1 - P_i(b)}{P_i(b)}} \quad (4.5)$$

となる。(4.5) 式の左辺は、消費者 i が製品 a を購入したか否かの情報が無い場合の製品 b の購買確率 $p_i(b)$ に対して、製品 a の購買の有無が分かった後の製品 b の購買確率 $p_i(b|a)$ は、(4.5) 式の右辺の比率だけ増減する事を示している。このため、製品 a の購買経験が将来の他の製品 b の選択に与える影響の強さを示していると解釈できる。影響の強さは、 $\sqrt{\left(\frac{1 - P_i(a)}{P_i(a)}\right)^{t_{i,a}}}$ に比例し、特に製品 a を購入した場合は $t_{i,a} = 1$ なので、 $\sqrt{\frac{1 - P_i(a)}{P_i(a)}}$ に比例して購買確率 $P_i(a)$ の小さい製品 a を購入したという情報の影響が大きくなる。逆に製品 a を購入しなかった場合は逆数の $\sqrt{\frac{P_i(a)}{1 - P_i(a)}}$ に比例し、購買確率 $P_i(a)$ が大きい製品を購入しなかったという情報の影響が大きくなる。

次に、消費者 i の全購買履歴 $\{u_{i,1}, \dots, u_{i,m}\}$ を条件とする製品 b の購買確率を推計するために、先に述べた通り次の 2 つの仮定を制約として設ける。

仮定 1: 交互作用統計量 $s_i(a, b)$ は消費者 i に依存しない。

仮定 2: m 個の製品の購買履歴を条件とする確率 $P_i(b|A)$ は、各 1 個の製品 a_j の購買の有無を条件とする m 個の確率 $P_i(b|a_j)$ の平均で推計できる。

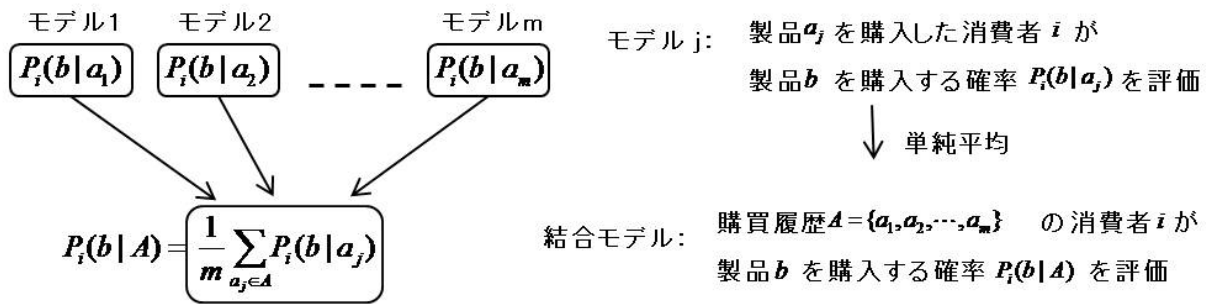


図 4-1. 条件付き確率推定モデルの結合

仮定 1 の制約は、表 4-2 における個人差は周辺確率 $P_i(a)$ と $P_i(b)$ だけで、周辺確率で調整した同時確率には個人差が無く、表 4-2 の個人差の自由度が 3 から 2 に減る事に相当する。この設定は、交互作用統計量を推計可能とするための制約であるが、製品カテゴリーの類似

関係が、製品カテゴリー知識によって形成されると解釈すると、製品カテゴリーの類似関係に個人差が無いという制約の是非は慎重に評価する必要がある、今後の課題とする。以降、消費者 i に依存しない交互作用統計量を $s_i(a, b)$ に代えて $s(a, b)$ と表記する。また仮定 2 の制約は、図 4-1 の通りモデルの結合の簡易計算に関する仮定である。

以上の 2 つの仮定を (4.4) 式に適用すると、提案モデルの算式となる次式 (4.6) を得る。

$$\begin{aligned}
 P_i(b|A) &= \frac{1}{m} \sum_{j=1}^m \left[P_i(b) \left\{ 1 + t_{i,a_j} s(a_j, b) \sqrt{\left(\frac{1 - P_i(a_j)}{P_i(a_j)} \right)^{t_{i,a_j}} \frac{1 - P_i(b)}{P_i(b)}} \right\} \right] \\
 &= P_i(b) + \frac{\sqrt{p_i(b)(1 - p_i(b))}}{m} \sum_{j=1}^m \left\{ t_{i,a_j} s(a_j, b) \sqrt{\left(\frac{1 - P_i(a_j)}{P_i(a_j)} \right)^{t_{i,a_j}}} \right\} \quad (4.6)
 \end{aligned}$$

(4.6) 式の右辺を計算するには、交互作用統計量 $s(a, b)$ と、周辺確率 $P_i(a)$ の値を知る必要がある。このうち交互作用統計量については、消費者を区別しないで表 4-2 の各事象の度数を数える事で、直接観測できる。一方、周辺確率 $P_i(a)$ の値は直接観測できない変数となる。

4.3.3 周辺確率の推定法 (1) - 定数と仮定

周辺確率 $P_i(a)$ の値を推定については、消費者 i の製品 a の購買確率 $p_i(a)$ が、過去の購買履歴とは独立では無いという設定を置くと、平均値の推定等における同一の確率分布に従う事象からの独立なランダムサンプリングという前提を置く事ができず、 $p_i(a)$ を推定する事が困難となる。このため本研究では、2 つの極端な周辺確率 $P_i(a)$ の推定方法を試し、その結果得られる条件付き確率 $P_i(b|A)$ の妥当性の評価を通して、間接的に好ましい周辺確率 $P_i(a)$ の推定法を評価する。

まず第 1 の推定法として、 $P_i(a)$ を定数とする。すなわち、購買履歴の無い初期状態では全ての消費者の全ての製品に対する購買確率が等しいと設定する。周辺確率 c (定数) を (4.6)

式に代入すると、

$$\begin{aligned}
P_i(b|A) &= c + \frac{\sqrt{c(1-c)}}{m} \left\{ \sqrt{\frac{1-c}{c}} \sum_{a_j \in A_i} s_i(a, b) - \sqrt{\frac{c}{1-c}} \sum_{a_j \in \bar{A}_i} s_i(a, b) \right\} \\
&\quad \text{ただし、} A_i \text{は消費者 } i \text{ が購買した製品の集合で、} \bar{A}_i = A - A_i \\
&= c + \frac{1-c}{m} \sum_{a_j \in A_i} s(a_j, b) - \frac{c}{m} \sum_{a_j \in \bar{A}_i} s(a_j, b) \\
&\quad c \text{ が } 1 \text{ に対して十分に小さいとして} \\
&\doteq \frac{1}{m} \sum_{a_j \in A_i} s(a_j, b) \tag{4.7}
\end{aligned}$$

(4.7) 式の $s(a_j, b)$ を類似係数で置き換えると、(4.7) 式は筆者の知る最も簡易なヒュリスティックベース法の協調フィルタリングの算式に一致する。(4.7) 式を行列表記すると、次の通り簡易な表現となる。

$$P = U \cdot S \tag{4.8}$$

ただし、 P は消費者 i の a_j 製品に対する購買確率 $P_i(a_j|A)$ を (i, j) 要素に持つ行列、 $U = \{u_{i,j}\}$ は 2.2 節で定義した通り消費者 i の a_j 製品に対する購買の有無 (0 または 1) を要素に持つ行列、 S は類似係数 $s(a_i, a_j)$ を要素に持つ行列とする。(4.8) 式は、製品間の類似係数 S と購買履歴 U による将来の購買確率 P の推定式と解釈できる。類似係数 S は購買履歴 U を用いて計算するため、過去の購買 U が将来の購買 P を説明する時系列の漸化式との解釈もできる。

4.3.4 周辺確率の推定法 (2) - 消費者の購買数と製品の販売数が独立と仮定

周辺確率 $P_i(a)$ の第 2 の推定法では、周辺確率 $P_i(a)$ は購買履歴 U とは関係が無いとする先の推定法とは対照的に、購買履歴 $U = \{u_{i,j}\}$ の周辺度数により $P_i(a)$ を推定する。全購買数を $N = \sum_{i,j} u_{i,j}$ 、消費者 i の各製品に対する購買確率の和を $P(i) = \sum_j u_{i,j}/N$ 、製品 a_j の各消費者による購買確率の和を $P(a_j) = \sum_i u_{i,j}/N$ と計算し、 $P(i)$ と $P(a_j)$ は独立であると仮定して、

$$P_i(a_j) = \frac{P(i) \times P(a_j)}{r} \tag{4.9}$$

と計算する。ただし、 r は観測期間と予測期間の差異に対応するための、(4.9) 式の左辺と右辺の水準を調整するスカラー量とする。購買履歴 U は過去の観測値なので、その行和と列

和が将来の購買予想確率の和である $P(i)$ および $P(a)$ に一致する必要は無いが、過去と将来の購買量には比例関係があると仮定して水準のみを r で調整する。

このとき、水準調整パラメータ r は周辺確率 $P_i(a)$ の水準を決定し、(4.5) 式の右辺の $\sqrt{\left(\frac{1-P_i(a)}{P_i(a)}\right)^{t_{i,a}}}$ の値を決める。図 4-2 の通り、この項は購買の有無 ($t_{i,a} = 1$ または 0) により確率 0 または 1 近辺で変化率が大きくなる。全体の確率水準が 0 に近づくと購買したという情報を重視し、逆に確率水準が 1 に近づくと購買しなかったという情報が重視される。よって、パラメータ r は購買経験の有無のウェイトを左右する働きをする。そこで、パラメータ r の推定の基準として、条件付き確率 $P_i(b|A)$ の予測力に関する検定統計量を最大化する r を求める。

周辺確率を定数とする推定法の (4.7) 式では、購買しなかった経験は将来の消費者の選好に影響を与えない算式となっている。このため、パラメータ r の最適値が存在すれば、周辺確率を定数とする推定法に否定的な結果となる。

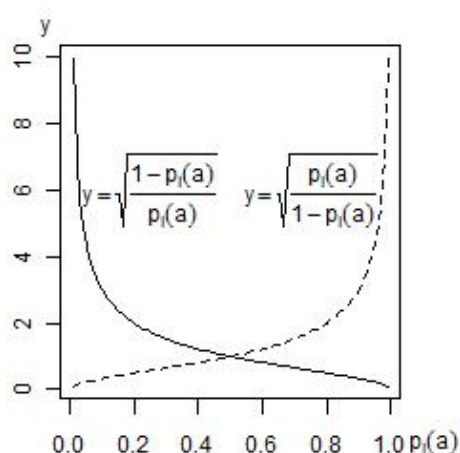


図 4-2. 周辺確率の水準と購買の有無別の重み関数

4.3.5 結果の規格化

次に、(4.6) 式で推定する条件付き確率 $P_i(b|A)$ の周辺確率について考える。条件付き確率 $P_i(b|A)$ に従った購買により購買履歴 U が得られるならば、 $P_i(b|A)$ の消費者 i についての集計は、消費者 i の購買数の期待値を示し、購買履歴 U の消費者 i についての集計と一致する。また製品についての集計も同様となる。実際には、消費者 i は購買の経験を通して条

件付き確率 $P_i(b|A)$ の値を変化させるので、2つのデータの消費者や製品に関する集計が一致する必要は無い。しかし $P_i(b|A)$ の変化が小さければ、消費者および製品に関する2つの集計値は近い値となる。そこで、(4.6)式により $P_i(b|A)$ を計算した後に、 $P_i(b|A)$ の消費者や製品に関する集計値が購買履歴 U の行和と列和に一致するように比例配分する規格化の処理を行う。

具体的な計算処理の手順は、まず各消費者の $P(i)$ について $P_i^1(b|A) = P(i) \frac{P_i(b|A)}{\sum_i P_i(b|A)}$ 、次に各製品の $P(a)$ について $P_i^2(b|A) = P(a) \frac{P_i^1(b|A)}{\sum_j P_j^1(b|A)}$ と計算し、 $P_i(b|A)$ を $P_i^2(b|A)$ の値で置き換える。ただし $P(i)$ と $P(a)$ は、(4.9)式における周辺確率の計算値の通りとする。

表 4-3. 周辺確率の推定に関する6通りの計算法

		推定結果の規格化	
		しない	する
周辺確率 の推定方法	Jaccard 係数の単純加算	jaccard(1)	jaccard(2)
	基準化交互作用統計量の単純加算	interaction(1)	interaction(2)
	基準化交互作用統計量の加重加算	weighted(1)	weighted(2)

以上の条件付き確率 $P_i(b|A)$ の推定法のパターンをまとめると、表 4-3 の通りになる。表 4-3 の各行は周辺確率の推定法に対応し、列は前述の通り推定結果の規格化の有無に対応している。第1行では、周辺確率を定数と置いた場合の (4.7) 式の $s(a, b)$ に、代表的な類似係数である Jaccard 係数を用いる (Gordon, 1999, pp.18)。Jaccard 係数を主効果 x (ただし行と列の主効果を等しいとする) と交互作用統計量 s の関数で表すと $\frac{2}{1+(1-s)(1-x)} - 1$ となり、Jaccard 係数は主効果 x と交互作用統計量 s に対して対称な感応度を持つ類似係数である事が分かっている (石田・西尾・椿; 2011)。

第2行では、(4.7)式の本来の計算通り (4.1) 式の交互作用統計量を用いる。また周辺確率の算式に (4.9) 式を用いた場合が、第3行に相当する。それぞれ、本論で参照するために「jaccard」「interaction」「weighted」と便宜的にラベル付けをする。また推定結果の規格化の有無は (1) あるいは (2) とラベルに追記する。下表 4-4 に購買確率の評価のための変数、パラメータ一覧を整理した。

表 4-4. 購買確率の推定のための変数

記号	意味	値の評価法
$u_{i,j}$	消費者 i の製品 j の購買の有無 (0 or 1)	購買履歴より判定
N	全購買数	$\sum_{i,j} u_{i,j}$
r	水準調整パラメータ	実験して最適値を選択 (図 4-3 参照)
$P(i)$	消費者 i の各製品に対する購買確率の和	$\sum_j u_{i,j}/N$
$P(a_j)$	製品 a_j の各消費者による購買確率の和	$\sum_i u_{i,j}/N$
$p_i(b)$	消費者 i の製品 b の購買確率	$\frac{P(i) \times P(b)}{r}$ (4.9) 式
$p_i(a_j)$	消費者 i の製品 a_j の購買確率	$\frac{P(i) \times P(a_j)}{r}$ (4.9) 式
$s(a_j, b)$	製品 a_j と製品 b の交互作用統計量	購買履歴より推定
t_{i,a_j}	消費者 i の製品 a_j の購買の有無 (-1 or 1)	購買履歴より判定

4.4 音楽 C D 販売 P O S データを用いたモデルの検証

4.4.1 目的と手順

本節では、前節で提案したモデルの有効性を検証する。検証データには音楽 C D 販売の I D 付き P O S データを用い、購買予測力をコルモゴロフ-スミルノフ検定で評価する。検証の概要は次の通りである。まず学習用と検証用のデータセットを 7 通り作成する。各データセットについて、学習データから交互作用統計量と周辺確率を表 4-3 の計算パターンに応じて推計し、(4.6) 式か (4.7) 式を用いて購買予測確率を求める。この値を、検証データにおける購買の有無で 2 グループに分け、2 グループの購買予測確率の分布の一致をコルモゴロフ-スミルノフ検定の片側検定で評価する。ただし学習データで購買のあったデータについては、繰返し購買の可能性が小さいので予め予測対象から除く。

コルモゴロフ-スミルノフ検定は、2 標本の母集団の確率分布が等しいという帰無仮説を検証するので、片側検定により仮説が棄却される事は購買がある場合の予測確率がそうでない場合に比べて大きい事を意味する。また検証の仕組みとして 2 つの標本の経験累積分布関数 (横軸の値よりも小さい標本数の比率を縦軸とする曲線) の差の最大値を評価するので、

2 標本の確率分布の形状や位置の差に依存しないノンパラメトリックな検定手法である。予測確率の分布を特定して検定する根拠に乏しい点を考慮し、ノンパラメトリックなコルモゴロフ-スミルノフ検定を用いた検証が適切と考えた。

具体的な検証手順として、最初に購買予測の重要なパラメータとなる (4.6) 式の水準調整パラメータ r を評価する。(4.6) 式を用いるのは表 4-3 の「weighted」の推定パターンの場合なので、「weighted(1)」の場合についてモデルの説明力が最大となる水準調整パラメータ r を評価する。まず網羅的に r の値を 10^l (ただし $l = 0.0, 0.2, \dots, 5.0$) と変化させ、それぞれについて周辺確率 $P_i(a)$ を (4.9) 式で求め、これを (4.6) 式に代入して条件付き購買確率 $P_i(b|A)$ を求める。次に $P_i(b|A)$ の値を検証データにおける購買の有無で 2 標本に分類し、コルモゴロフ-スミルノフ検定で 2 つの確率分布の差異を評価する。最適な水準調整パラメータ r は、7 通りのデータセットに共通な値として 1 つ決める。

条件付き購買確率 $P_i(b|A)$ の有効性の検証は、表 4-3 の 6 通りの各推定法について行う他、2 つの推定法の有効性の比較も行う。比較の手順は、最初にそれぞれの推定法で $P_i(b|A)$ を計算し、これを順位統計量に変換し、検証データで購買のあった順位統計量を抜き出して、2 通りの順位統計量の分布の一致をコルモゴロフ-スミルノフ検定の両側検定で評価する。学習データで購買のあったデータを予め除いておくのは先程と同じである。

4.4.2 データ

大手セル・レンタル CD チェーン店から個人が特定できないように加工した上で提供頂いた、音楽 CD 販売の購買履歴 417,380 件の ID 付き購買 POS データを検証に用いる。購買の期間は 2002 年 11 月 1 日から 2003 年 12 月 21 日までの 1 年余りで、東京と大阪エリアの計 5 店の購買履歴が対象となっている。検証では、過去 6 ヶ月間の購買履歴を学習データに用いて翌月の購買を予測し、翌月の購買履歴を検証用データに用いて予測の一致を評価する。例えば第 1 期は 2002 年 11 月 1 日から 2003 年 4 月末までの購買履歴を学習データとして 2003 年 5 月の購買を予測する。第 1 期以降は 1 カ月ずらしてこの手順を繰り返し計 7 期の検証を行う。また交互作用統計量や周辺確率の推計が満足にできない評価対象を除く目的から、学習期間に月間平均 1 枚以上の製品を購入した消費者と、月間平均 2 枚以上売れた製品に限定する。

本論で実証分析対象とする音楽 CD は、衝動買いが少なく、事前に購入製品を決めてか

ら店舗に行く買い方が全体の82.5%を占めるという調査結果があるように（日本レコード協会, 2009）、探索コストが大きいという特徴を持っている。特に社会人に比べて時間的余裕のある10代の消費者は、幾度も聴いて自分が好きな楽曲やアーティストを評価してから購買を決める傾向がある（日本レコード協会, 2008）。このような特徴を踏まえて、音楽CDのID付きPOSデータの先行研究（オペレーションズ・リサーチ, 2007, vol.52, no.2, データ解析コンペティション特集号; 勝又・阿部, 2007）では、音楽CDは嗜好の異質性が大きいと、顧客セグメントや個人単位の嗜好の推定が重要であると指摘している。

4.4.3 結果

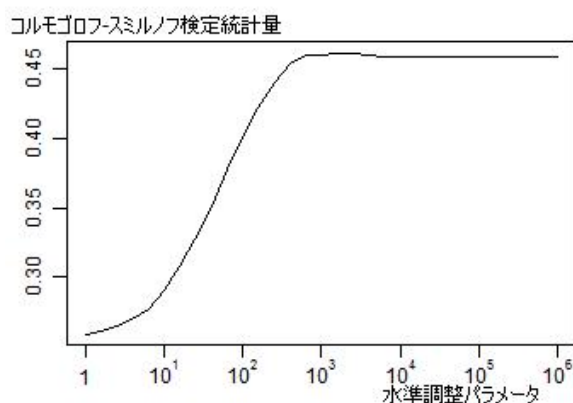


図 4-3. 音楽CD販売POSデータの
購買試行回数と検定統計量の関

係

図 4-3 に水準調整パラメータとコルモゴロフ-スミルノフ検定統計量の対応を示している。グラフの各点は7期の評価の平均値としている。図 4-3 の曲線は、水準調整パラメータが増えるにつれて右肩上がりとなり、10³ 近辺で最大となった後、若干下がって横ばいになっている。この結果から、最適な水準調整パラメータとして $r = 10^3$ を用いる。

図 4-3 の曲線の形状は、水準調整パラメータを大きくして、ある製品の購買があったという情報のウェイトを一定値まで高める必要があるが、それ以上大きくして製品の購買が無かったという情報のウェイトを0に近づけても、説明力の低下がほとんど無い事を示している。すなわち製品の購買が無かったという履歴には、ほとんど情報量が無いと解釈できる。模擬のPOSデータを用いた評価では、第4章末に Appendix として掲載した図 A4-1 の通

り、購買が無かったという履歴にも情報があるという結果となり、実証データと異なる結果を示した。この違いの要因として、特定の店舗のPOSデータには消費者が他の店舗で製品を購入したりレンタルした履歴が含まれず、購買履歴が無い事と消費者の購買行動の間にギャップがあるためと考えられる。

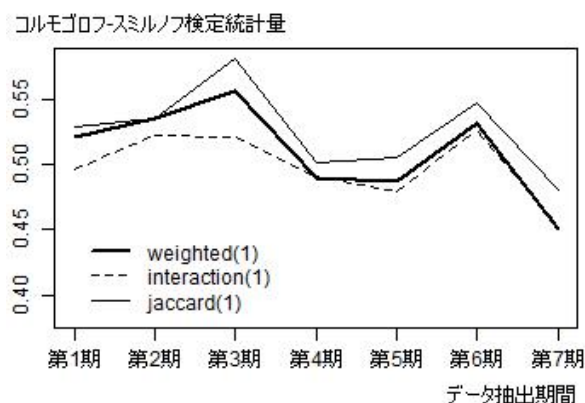


図 4-4-1. 周辺確率推定法の比較
音楽CD販売POSデータ・規格化なし

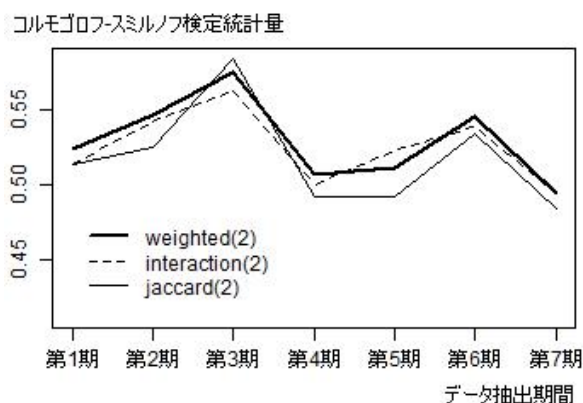


図 4-4-2. 周辺確率推定法の比較
音楽CD販売POSデータ・規格化あり

表 4-5. 音楽CD販売POSデータを用いた周辺確率の推定方法の比較

	帰無仮説	有意水準と該当回数 (全 7 回)		
		0.1%	1%	5%
周辺確率の推定法の比較 (両側検定、表 3 の行に対応)	jaccard(1) = interaction(1)	5 回	7 回	7 回
	interaction(1) = weighted(1)	5 回	6 回	6 回
	jaccard(2) = interaction(2)	0 回	1 回	2 回
	interaction(2) = weighted(2)	0 回	0 回	1 回
推定結果の規格化の有無の 比較 (片側検定、表 3 の列に対応)	jaccard(1) > jaccard(2)	0 回	0 回	0 回
	interaction(1) > interaction(2)	6 回	7 回	7 回
	weighted(1) > weighted(2)	0 回	1 回	3 回

図 4-4-1 に表 4-3 の第 1 列に相当する規格化しない推定法の検定結果を、図 4-4-2 に規格化する推定法の検定結果を図示している。これらの 6 通りの 7 期の全ての検定で 0.1%水準で有意となった。推定結果を規格化しない図 4-4-1 において、検定統計量の比較として「interaction(1)」<「weighted(1)」<「jaccard(1)」という優劣を見てとれる。推定結果の一致

を検定した表 4-5 から、0.1%水準で 7 回の中の 5 回で一致が棄却されている。推定結果を規格化した図 4-5-2 では、規格化前と比べて「jaccard」の説明力の改善が小さく、「interaction」の改善幅が最も大きい。

この理由は、交互作用統計量は jaccard 係数と違って周辺確率と独立な統計量であるため、周辺確率を考慮した規格化による予測力の改善効果が大きいと解釈できる。総じて若干ながら「weighted(2)」が最も安定して高い説明力を示している。これらの傾向は模擬の POS データの検定結果でより顕著となり、図 A4-2 と図 A4-3（第 4 章末に Appendix として掲載）において明確に確認できる。

4.5 類似関係を利用するプロモーションへの応用

4.5.1 売上の交差弾力性と評価手順

ある 2 製品の類似係数が高いという事は、それらの製品が共に購入され易い事を表している。そこで、ある製品の売上を伸ばすためには、(1) 2 製品の類似を顧客に強く認識させて補完の性質を強める、(2) 該当する製品と補完関係にある製品の売上を伸ばす、というプロモーションを考える事ができる。これらの類似関係を利用する手法は、間接的ではあるものの、単独のプロモーション効果に加えて波及効果や累積効果を期待できる。また、補完関係による購買は既存顧客に重ね買いを促すため、新規顧客を獲得するのに比べて低コストのプロモーションで売上増加が見込める。

類似関係を利用するプロモーションを考える際に、音楽業界の商習慣上の特徴であるレコードレーベルの役割を考察する必要がある。レコードレーベルは、レコード会社やレコード会社の部門名に相当する組織で、元はレコードに張られたラベルを意味している。一般にアーティストは、一つのレコードレーベルに所属し、レコード会社は音楽性が似ているアーティストを同じレコードレーベルにグループ化する。これにより、音楽アーティストやその CD 製品は、同じ企業ブランドを持つブランド・ポートフォリオ (Aaker, 2004) を構成していると思える。このような商習慣が定着している理由として、音楽業界には選択と集中の経営戦略が適していると解釈できる。例えば、家電市場では消費者は掃除機も TV も購入するため、製品ラインを多角化し、それぞれにおいて高品位な製品を同一ブランドの下で提供する戦略が考えられる。一方で、音楽 CD 市場の場合は、消費者の嗜好す

る製品がクラシックやアイドルなどの特定の製品カテゴリーに限定され易い特徴がある。消費者の関与も一部の製品カテゴリーに集中し、それらの製品に関して豊富な知識を持つ消費者も多い。このため、選択と集中が一つの望ましい戦略となる。Aaker(2001)は、集中化戦略について、「より大きな競合相手の資産および能力を回避し、ポジショニング戦略を提供して競合他社の圧力を弱める。」利点があるとしている。集中化戦略は市場を細分化して特化することで競争優位に立つ戦略であり、企業間提携による資産や能力の補完と異なるベクトルの戦略である。レコード会社がレコードレーベルという小組織・事業部門を作り独立性を持たせるのも、集中化戦略では組織のスケールメリットが小さいからと解釈できる。このため、音楽CD市場の場合は、異なる企業間でシナジー効果を享受する提携戦略をとらないと考えられる。類似関係を利用したプロモーションを行う際も、他社との提携を志向することなく、同じレコードレーベルの製品（アーティスト）同士での類似関係による売上のシナジー効果を求めると考えられる。

アーティスト達をプロモーションする組織が同じなら、同じメディアやイベントに所属アーティストを露出し易く、類似関係を伝え易い。例えばレコード会社アップフロントワックスに所属するアーティストが、テレビ番組のハロー!モーニング（テレビ東京で2000年4月～2007年4月放送）に出演しているのを観た視聴者は、特に説明が無くても「モーニング娘。」に代表されるハロー!プロジェクトのメンバーと認識したであろう。音楽性の似通ったアーティストは競合する可能性も高いと思われるが、同じレコードレーベルに所属していれば、新譜の発売時期をずらすなど、競合要因を打ち消し易い。そして、あるアーティストの顧客層の拡大が、同じレコード会社の別のアーティストの顧客層の拡大につながるシナジー効果を享受できる利点がある。

本節では、前節で検証したモデルを用いて類似によるプロモーション効果の高いアーティストの組と、所属レコードレーベルとの関係について分析する。評価するデータには、モデルの検証に用いた音楽CD販売のID付きPOSデータの最初の期間のデータセットを用いる。アーティストの関連の強さは、需要の交差価格弾力性に倣い、売上の交差弾力性 $\delta_{A,B}$ を次の通り定義して利用する。

$$\text{売上の交差弾力性: } \delta_{A,B} = \frac{\text{アーティスト B の売上数量の変化率}}{\text{アーティスト A の売上数量の変化率}} \quad (4.10)$$

上式は、需要の交差価格弾力性の場合の分母の価格変化を売上数量の変化に置き換え、便宜的にマイナスの符号を除いた算式になっている。

アーティストAの売上げが増える事による、アーティストBの売上の増分の評価は、次の通り行う。まず、モデルの検証と同様に消費者の購買確率を評価し、アーティストAの購買確率が高い消費者から順に実際の売上げの10%に相当する消費者が新たに購買すると仮定し、これらの消費者の模擬の購買履歴をPOSデータに追加する。仮定する売上増の水準について、試行的な数値計算を行い、アーティストAの新たな購買を仮定する消費者の比率を小さくすると、集計対象となる人数が少なくなり、新たにアーティストAを購買する消費者に誰が選ばれるかといった偶然の影響が大きくなるが、比率を10%に設定すると、安定的な結果となる事を確認した。購買を仮定する消費者の比率をさらに増やすと、今度はアーティストAの潜在顧客として見込みの薄い消費者の購買を仮定することになり、実際的な設定から外れる。そこで、安定した結果が得られ、売上増加の目標として現実的な値として10%の水準を設定した。本実験では、アーティストAの新たな購買の発生による、アーティストBの条件付き購買確率の増分の総計を、補完関係による購買の効果として評価する。このように類似関係による売上の増分を評価できるのは、提案モデルが確率モデルである事の利点である。

4.5.2 評価結果

表 4-6. 関連購買の大きいアーティストの上位 10 組

No	アーティスト名				所属レコード会社名	
	売上 順位	要因	売上 順位	結果	要因アーティスト所属	結果アーティスト所属
1	25	松浦亜弥	79	後藤真希	アップフロントワークス	アップフロントワークス
2	2	浜崎あゆみ	6	BoA	avex	avex
3	23	KinKi Kids	90	J-FRIENDS	ジャニーズ・エンタテイメント	(ジャニーズ)
4	2	浜崎あゆみ	13	Every Little Thing	avex	avex
5	25	松浦亜弥	147	ごまっとう(後藤真希、松浦亜弥、藤本美貴)	アップフロントワークス	アップフロントワークス
6	1	CHENISTRY	9	平井堅	ソニー・ミュージックエンタテインメント(日本)	ソニー・ミュージックエンタテインメント(日本)
7	25	松浦亜弥	74	藤本美貴	アップフロントワークス	アップフロントワークス
8	2	浜崎あゆみ	3	B'z	avex	ビーイング
9	79	後藤真希	147	ごまっとう(後藤真希、松浦亜弥、藤本美貴)	アップフロントワークス	アップフロントワークス
10	38	モーニング娘。	79	後藤真希	アップフロントワークス	アップフロントワークス

表 4-6 に期間併売され易く売上の増加数量が高いと評価したアーティスト上位 10 組を示

した。表の売上順位は、売上増を見込む前の実際の売上順位を示している。第8位の「浜崎あゆみ」から「B'z」への影響の組み合わせ以外の9組は、同じレコード会社に所属し、まるでレコード会社へのロイヤルティーが働いているかのようである。ベースとなる実際の売上数量が多い方が期間併売として購買による増分も多くなり易いので、元の売上の大きいアーティストが多いが、「モーニング娘。」に代表されるレコード会社アップフロントワークスのメンバー同士は、元の売上げ順位が低いにも関わらず上位にランキングされている。このため、これらのアーティストを同時顧客を多く獲得している事が分かる。

表 4-7. 市場全体への影響が大きい上位 10 位のアーティスト

No.	売上順位	要因アーティスト名	所属レコード会社名
1	2	浜崎あゆみ	avex
2	6	BoA	avex
3	16	TLC	Sony Music Entertainment
4	13	Every Little Thing	avex
5	1	CHEMISTRY	ソニー・ミュージックエンタテインメント (日本)
6	11	宇多田ヒカル	EMI ミュージック・ジャパン
7	25	松浦亜弥	アップフロントワークス
8	22	ジェニファー・ロペス	Sony Music Entertainment
9	26	中島美嘉	ソニー・ミュージックエンタテインメント (日本)
10	44	175R	EMI ミュージック・ジャパン

次に関連購買による売上増加量を全てのアーティストについて合計し、表 4-7 に市場の売上へ貢献が大きいアーティスト上位 10 位をまとめた。表 4-6 と比較すると、限られた消費者に訴求する女性アイドルがランキング外になっている。これらのアーティストは、類似関係の影響の強さと潜在顧客の多さという広がりバランスがとれているので、マスメディアを通して消費者にプロモーションするのに適しているアーティストと思われる。一方、アイドル性の強いアーティストは、強い嗜好を持つ少数の消費者に好まれる傾向があるので、ターゲティングを定めてプロモーションするのに適していると思われる。

音楽CD市場では、レコードレーベルによる選択と集中戦略が観察されるため、同じレコードレーベル内のアーティストの類似関係が高く、提案した購買確率の推定式の解釈として、売上の相乗効果があると推測した。推測通りの結果を得た事は、提案した推定式の妥当性を示す内容であると考えられる。

4.6 結論と今後の課題

4.6.1 結論

全ての消費者に同じ交互作用統計量を適用するという提案モデルの制約を観念的に述べると、全ての消費者の市場観が同じという事になる。そして提案モデルが有効であるという結果は、この制約が消費者行動において自然なものである事を示唆している。そして、2つの制約から理論的に導出したモデルの算式は、協調フィルタリングのアルゴリズムと一致する事を示せた。

モデルの検証結果について、まず水準調整パラメータには最適値が比較的安定して存在する事が判かる（図 4-3 と図 A4-1(第 4 章末に Appendix として掲載)）。この事は、購買した事と同様に購買しなかった事も情報として意味があり、情報の重みとして最適なバランスがあると解釈できる。モデルの有効性については、表 4-3 のパラメータ推定のどの方法についても 0.1%水準で有意となる事を確認できる。推定法の比較としては、購買の有無の情報を加重加算して規格化する方法が優位となったが、より単純な計算として交互作用統計量を単純加算して規格化しても同様に良い説明力を得る事ができる。また、モデルを用いて類似関係による購買の高いアーティストの組を評価したところ、上位 10 組の中で 9 組が同じレコード会社に所属している事が分かる。このため同じレコード会社によるプロモーションには、補完関係による売上増加効果があると解釈できる。

提案モデルは、音楽CD市場に限らず、製品評価が経験に影響され易い財の評価に、特に有効な手法であると考ええる。そしてPOSシステムなど購買履歴を蓄積する仕組みがある企業においては、消費者や製品属性に関する調査コストを必要せず、適宜利用可能な利便性の高いモデルである。

4.6.2 今後の課題

提案モデルの発展として、モデルの制約を緩和する方向と、時系列情報などの変数を新たに組み入れて精緻化する2つの方向がある。まず緩和の方向について、第1の制約である製品の類似関係の認識の共通性に対して、消費者の異質性を許容する事が考えられる。例えば、製品に対する知識の差異や、新製品の購買タイミングで特徴付けられる革新者・追随者の消費者セグメントによって、製品の類似関係の認識やタイミングが異なっているとすれ

ば、これらの消費者セグメント毎に類似関係を設定するのが適切であろう。また第2の制約について、2製品の購買に関する同時確率から得た購買予測確率を、単純平均して複数製品の期間併売とする計算法が簡便過ぎる懸念がある。また新たな変数を組み入れる方向として、製品の属性の類似関係を利用する手法がある（石田・西尾・佐藤, 2008）。これにより、購買履歴の無い新製品には協調フィルタリングが使えないという限界（Ansari, Essegaiier & Kohli, 2000(1)）を緩和できる。そこで、複数の類似関係を組み合わせて評価するモデルの仕組みを構築する事が今後の課題となる。

レコード会社のレコードレーベルを用いたプロモーションについては、アーティストの類似性がレコード会社のプロモーションによって高められたのか、あるいはレコード会社が類似性の高いアーティストを集めた結果なのかについて、議論できていない点が課題である。また、レコード会社が同じであれば、プロデューサーなどのスタッフが同じになり易いため、意図されたプロモーションとは関係無くアーティストの類似性が高まるという効果もある。類似関係の認識要因を整理し、具体的なプロモーション手法を評価する事が今後の課題である。

第4章の Appendix: 模擬データを用いた検証

第4章3節における提案モデルの検証を、実データに代えて模擬POSデータで行う。

模擬POSデータの作成

下記の手順を繰り返し、計100セットの模擬のPOSデータを生成する。

1. 配置 m 人の消費者と n 個の製品を d 次元のユークリッド空間に配置する。各消費者と製品の位置は、各軸でそれぞれ独立に $[-1, 1]$ の範囲の一樣乱数を d 個発生して各座標値とする。
2. 購買判定 各消費者について平均 h の正規乱数で半径 r_1 を生成し、これを該当する消費者の購買可能半径とする。同様に各製品について平均 $1 - h$ の正規乱数で半径 r_2 を生成し、この製品の購買可能半径とする。そして消費者と製品の組合せについて、お互いの半径の合計 $r_1 + r_2$ と座標の距離 L を比較して $L < r_1 + r_2$ の時に限り、平均値が $p_{max}(1 - \frac{L}{r_1 + r_2})$ の2項分布で購買の有無を決定する。

3. 学習と検証データ 先のステップで判定した各消費者と製品の購買の組合せを、確率 $q : 1 - q$ で検証データと学習データに割り振る。

模擬POSデータの作成方針として、周辺確率と交互作用統計量の分布が、実証に用いる音楽CDの販売データと一致するようにパラメータを設定する。試行を経て経験的に 1. 消費者数 $m = 20,000$ 、2. 製品の数 $n = 500$ 、3. 空間の次元数 $d = 5$ 、4. 消費者の購買可能距離平均 $h = 0.4$ 、5. 最大確率 $p_{max} = 0.5$ 、5. 検証データ比率 $q = 20\%$ の各パラメータを採用する。

結果

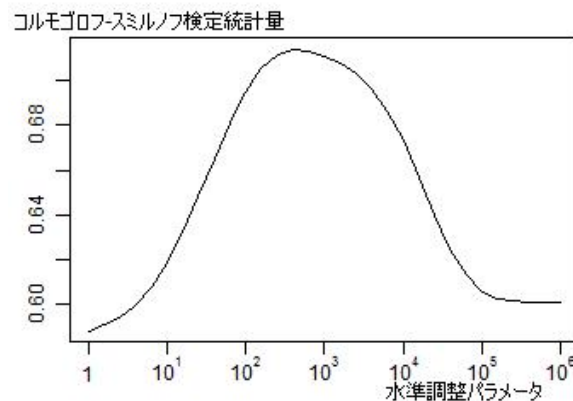


図 A4-1. 模擬 POS データの水準調整パラメータと検定統計量の関係

100 個の内の 58 個のデータセットで $r = 10^{2.6}$ の時に検定統計量が最大となる。

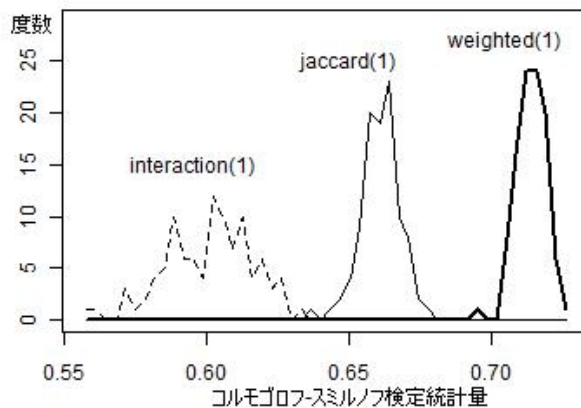


図 A4-2. 周辺確率推定法の比較
模擬POSデータ・規格化なし

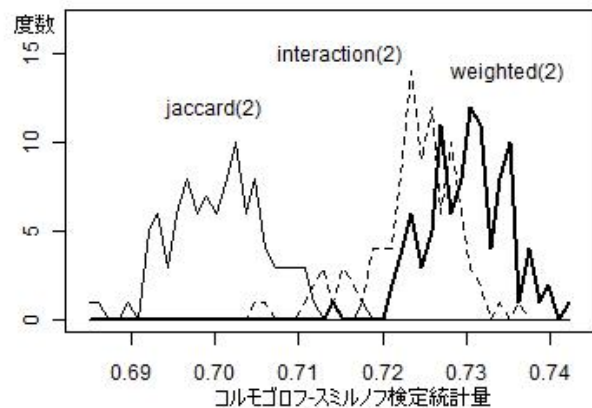


図 A4-3. 周辺確率推定法の比較
模擬POSデータ・規格化あり

表 4-7. 市場全体への影響が大きい上位 10 位のアーティスト

No.	売上順位	要因アーティスト名	所属レコード会社名
1	2	浜崎あゆみ	avex
2	6	BoA	avex
3	16	TLC	Sony Music Entertainment
4	13	Every Little Thing	avex
5	1	CHEMISTRY	ソニー・ミュージックエンタテインメント (日本)
6	11	宇多田ヒカル	EMI ミュージック・ジャパン
7	25	松浦亜弥	アップフロントワークス
8	22	ジェニファー・ロペス	Sony Music Entertainment
9	26	中島美嘉	ソニー・ミュージックエンタテインメント (日本)
10	44	1 7 5 R	EMI ミュージック・ジャパン

水準調整パラメータの値は図 A-2-1 より $r = 10^{2.6}$ を採用する。図 A-2-2 と図 A-2-3 の全ての検定において 0.1%水準で有意となる。

第5章 消費者の選好構造の空間的表現の研究

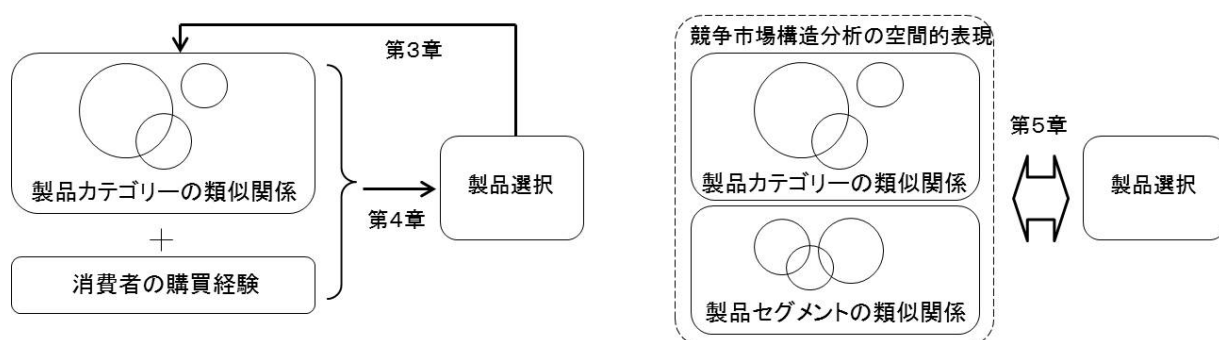


図 5-1. 第3章～第5章における製品カテゴリーの類似関係と製品選択の関係評価

5.1 はじめに

第3章では、消費者の製品選択から製品カテゴリーの類似関係を推定する手続きの妥当性を評価した。第4章では、逆の経路として、製品カテゴリーの類似関係から消費者の将来の製品選択を説明できるモデルを示した。残る課題は、競争市場構造分析の手法を踏まえて、製品カテゴリーに対応した消費者セグメントを行動面から推定し、製品カテゴリーに対する消費者の選好構造を空間的に表現する事である。

消費者セグメントを製品カテゴリーに対応させて推定するのは、例えば、廉価品と高額品という製品カテゴリーを設ける理由を、価格を重視する程度が異なる消費者セグメントが存在するためと考えるように、製品の競合と消費者の選好構造を、コインの表裏のように一体として同時に捉える競争市場構造分析の方針を踏まえている。このため、本章でも、先に推定した製品カテゴリーに依存して消費者セグメントを推定するのではなく、それぞれを同時に推定する。また、競争市場構造分析の表現には、ポジショニングマップなどの空間的表現と、離散的なラベルとしての分類といった非空間的表現があるが（第2章、表 2-2）、ID 付き POS データが消費者の時間的に連続な購買記録である利点を活かすには、市場構造の

推移を連続に捉えられる空間的表現が都合が良い。空間的に表現された市場の推移を観察することにより、消費者の購買行動に関する新たな知見を発見できる場合もあると考える。

消費者セグメントの空間的な表現は、個々の製品に対応付けられた多次元の直交座標軸を持ち、各軸の座標によって消費者の選好の強さが関連付けられた、消費者の配置とする。例えば市場に製品が10個あり、製品カテゴリーの類似として10個の製品が無関係（独立）の場合は10次元の空間表現とする。2製品に対する選好のみに関心がある場合は、該当する2製品に対する消費者の選好のみを取り出して、その選好の強さを各軸の座標の値とすれば、2製品への消費者の選好構造が2次元平面に示される事を意図している。しかし、射影された布置において製品間の類似や競合関係を表すには、製品間の類似関係を反映して軸を斜交するといった精緻化が課題として残っている。この点については、5.2.2. 研究の課題において詳述し考察する。

製品カテゴリーの類似関係と消費者セグメントの選好の類似関係を、空間的表現としないで類似係数行列の表現形式のままにしているのは、クラスター分析などの限られた統計手法しか適用できない。また、消費者と製品のそれぞれの類似係数行列を関係づけるのが難しい。これに対して、製品カテゴリーを座標軸に対応させ、消費者セグメントが配置に対応する空間的表現を提示できれば、様々な統計手法が適用可能となり、そのような空間表現は、競争市場構造分析を行う際の汎用的な分析フレームとして利用し易くなる。また、統計手法を使わずとも、競合関係に変化が見られる2製品を評価軸として、潜在顧客や購入顧客の配置の推移を視覚的に観察するといった分析のアプローチもできよう。

第5章では、まず、提案する空間的表現が消費者の選好構造を表している事を確認する。次に、提案する空間的表現が実務的に有用である証左として、新製品の売上推移に関する実証分析から新たな知見を導く。提案する空間的表現が消費者の選好構造を表しているならば、空間における消費者の分布が、新製品の売上推移と関係があって当然である。新製品の売上推移については普及モデル（Bass, 1969）による説明があるが、消費者の製品選択要因には普及理論が想定する直接コミュニケーション以外の要因も大きいと考える。消費者の選好構造としての消費者の分布の形状から、製品の売上推移パターンを予測し、プロモーション戦略の立案・判断材料としての実務的な有用性を示すことが実証分析の課題である。

5.2 先行研究と課題

5.2.1 先行研究

第5章の研究課題は、行動面から捉えた消費者の選好構造を、競争市場構造分析のアプローチにより空間的に表現する事である。また、そのような空間的表現の有用性を示すために、消費者の選好を表す空間的分布の形状と、製品の売上推移パターンとの関係を見出してプロモーション戦略に役立てることである。したがって、競争市場構造分析、空間的表現、消費者の異質性の扱い、普及理論について先行研究のレビューを行う。

表 2-2 (再掲) . 競争市場構造分析手法・モデルの類型

	空間的表現	非空間的表現
判断データ	製品マップなど、因子分析、多次元尺度構成法	Urban, Johnson & Hauser(1984) など
行動データ	Kamakura & Russell(1989) など	井上・中西(1990)、井上(1992) Grover & Srinivasan(1987,1989)、 Moe & Fader(2001)

井上(1992) 表1 を基にして作成。

競争市場構造分析の類型として、井上(1992) は表 2.2 (再掲) の通りデータの特徴と分析の表現のタイプによる4類型を示し、競争市場構造分析は顧客サイドから競争状態を把握することにより、マーケティング戦略の形成に関して有用な情報を提供しようとするものであると、位置付けている。行動データを用いた競争市場構造分析では、Kamakura & Russell(1989) などマーケットシェアやブランドスイッチにおける競争関係に着目するアイデアがある。Kamakura & Russell(1989) は、ブランドのシェアの変化の要因として、競合ブランドに対する競争力と、競合ブランドのプロモーションからの脆弱性の2軸の空間にブランドを配置する手法を提案している。しかし、書籍や音楽CD市場のように新製品が売上の大半を示める市場では、マーケットシェアが短期間に大きく変動し、リピート購買が期待できないためブランドスイッチにおける競争関係を測定しづらいという課題がある。また、類似係数を用いて消費者セグメントを捉える場合(Wedel & Kamakula, 2000)、類似係数は距離としての性質を備えていないため空間表現に適さないという課題があった。Moon & Russell(2005) は類似係数の行列を分解して低次元の線形空間の表現に近似することで、この課題に対応している。ところが、第3章の結果は、製品カテゴリーの類似関係を捉える

類似係数として交互作用統計量が適切であり、交互作用統計量は2項分布の相関係数として（広津, 1992）、距離としての性質を備えているため、交互作用統計量による評価は空間的表現にもなり得る事を示唆している。空間的な表現を持つ他の類似係数としては、Jaccard 係数と Simple matching の類似係数行列が半正定値となるため自然な距離が対応する事が知られている（Gower & Legendre, 1986; Gordon, 1999）。そして、第3章の結果から、Simple matching が有用性に乏しいクラスタリングを導いた事から、自然な距離が対応するために、類似係数行列が半正定値となる事は必要条件ではあるが、有益な空間的表現を導く十分条件では無い事が分かる。

消費者の選好構造の空間的表現は、同一製品市場で互いに競合している各ブランドを消費者がどのように理解しているかを多次元上に図示したプロダクトマップあるいは知覚マップ（片平, 1987）を意図した表現である。視覚的な図示は必要に応じて2次元平面等に射影すれば描画できるので、多次元線形空間上に各製品と消費者を配置することを目的とする。ただし、市場の特徴を捉える2次元平面の特定については、今後の課題とする。空間上の配置の意味の与え方には、ベクトルの関係で表現する理想ベクトルモデルと、2点の距離で捉える理想点モデルの、2つがある（阿部, 2003; 柳井他, 1990, 7章）。理想ベクトルモデルでは、ベクトルの方向の近さが選好の強さに対応し、理想点モデルでは2点間の距離が選好の強さを表す。例えば、異なる配置にある製品Bと製品Cと、その中点にある製品Aを考えた場合、製品Aから見て製品Bと製品Cの距離は同じであるが、ベクトルとしては反対方向にある。このように、空間的選好構造として点間距離で評価するか、ベクトルの内積として評価するかは大きな違いとなり、理想ベクトルモデルと理想点モデルのどちらを用いるかは、慎重に検討すべき事項となる。交互作用統計量から導いた配置の場合、交互作用統計量は2項分布の相関係数としてスケールが基準化されているため、製品や消費者の配置は原点から等距離に配置できる。このため、理想ベクトルモデルと理想点モデルにおける大きな差異が発生しない。これは、交互作用統計量が、多く買う消費者や良く売れる製品といった主効果の評価とは独立な類似関係を評価している事に相当している。

表 5-1. 消費者の選好構造の空間的表現が備えるべき性質

1. 製品カテゴリーに直交座標系が対応.
2. 消費者セグメントと製品カテゴリーの対応.
3. 消費者の座標と選好の強さの対応.

阿部 (2003) は理想ベクトルモデルと理想点モデルなどのジョイントスペースを比較し、空間的表現としての優劣を評価する 3 つの基準として、(A1) 空間的表現を与えるモデルの適合度 (対数尤度)、(A2) モデルによる製品選択の予測確率、(A3) ヒット率 (実際の正解率) の 3 点を挙げている。これを参考に、本研究において消費者の選好構造の空間的表現が備えるべき性質として、表 5-1 の通り、(B1) 製品カテゴリーに直交座標系を対応する、(B2) 消費者セグメントと製品カテゴリーが対応する、(B3) 消費者の座標と選好の強さが対応する、の 3 点とする。一般にプロダクトマップの描画では、製品属性など変量に対する因子分析や多次元尺度構成法などのモデルを用いるが、行動面のデータを用いる本研究ではモデルを用いることなく個々の製品を各軸に対応させる。モデルを用いないため、阿部 (2003) の基準のモデルの適合度や予測確率に対応する基準がないが、その代わりにモデルにおいて実現すべき、空間の座標系の構成 (B1) と競争市場構造分析としての消費者セグメントと製品カテゴリーの対応 (B2) の要件の確認を必要とする。また、(B3) 消費者の座標と選好強さの対応は (A3) と対応している。

表 2-3 (再掲) . モデルにおける異質性の所在

	異質性の所在	異質性なし	異質性あり
製品カテゴリーと製品選択の関係	製品カテゴリーの類似関係	製品カテゴリーの類似関係は消費者に共通	製品カテゴリーの類似関係は消費者により異なる
空間的表現	軸の構成	全消費者に 1 枚のポジショニングマップ、消費者の配置が異質	消費者毎に異なるポジショニングマップ、評価軸や製品の座標配置が異質
因子分析	因子負荷量	一般の因子分析、因子負荷量が共通で、因子得点が消費者で異質	因子負荷量が異質 Ansari & Jedidi(2002)
共分散構造分析	因子構成と付加量	一般の共分散構造分析	グループで因子構成が異なる 多母集団分析

消費者の異質性の扱いについて、第 2 章で先行研究をレビューした通り (表 2-3, 再掲)、軸の構成に対応する製品カテゴリーの類似関係に消費者の異質性が認められれば、空間的表現の軸の構成に異質性を与えるべきである。第 4 章の議論および結果は、行動面のデータが

ら製品カテゴリーの類似関係における異質性を評価するのは、データの自由度の観点から難しく、また、製品カテゴリーの類似関係に異質性が無いと仮定したモデルが消費者の将来の製品選択を優位に予測しており、意味のある結果を示した。この結果を踏まえ、第5章においてもモデルとして消費者の異質性を扱わない設定とし、因子分析によるプロダクトマップの描画と同様に、分析者が消費者の布置に意味を見出す表現とする。

新製品の普及モデル (Bass, 1969) は、市場における消費者の近さと時間の連続的な変化を扱うモデルである。普及モデルでは新製品の普及を革新者と模倣者の2セグメントの消費者間の接触によって説明し、直接コミュニケーションや接触の起き易さを扱っている。また、Rogers (1983) は採択者のプロフィールと採択時期を関連付けて、普及段階を5段階 (革新者、初期採択者、前期追従者、後期追従者、遅延者) に分類している。これらの初期の普及過程の研究では、自動車やTVなどの耐久財の数十年単位のライフサイクルのプロセスが研究対象とされたが、続いて普及理論を映画や音楽などの短期のライフサイクルに拡張する研究へと発展し (Krider & Weinberg, 1998; Sawhney & Eliashberg, 1996; Lehmann & Weinberg, 2000; 荒木, 2002)、普及理論は様々なライフサイクルの製品に適用可能である事が知られている。Adomavicius & Tuzhilin (2005) は、普及理論を一般化して流行現象として捉える概念モデルを提案している。また、普及理論に近い研究領域として、直接コミュニケーションの影響を扱う口コミの研究がある。濱岡 (1993) は、この研究分野のサーベイを行っており、言動が周囲に与える影響の大きいオピニオン・リーダーの存在や、相乗的な波及効果など、マスメディアとは異なる伝達のメカニズムの解明が主要な研究課題とされると述べている。普及モデルも口コミの知見も、ともに新規採択者は空間的に伝播して広がり、やがて潜在的な採択者への浸透が終了し、ライフサイクルを終えるという考え方に基づいている。この知見を類似関係から導かれる競争空間に援用すると、新製品の革新的消費者が距離の近い追従的消費者に接触することで、新規採択者の空間的配置は次第に空間全体に拡散して広がる事になる。空間全体に広がった消費者間のお互いの距離は遠くなり、消費者間の類似は小さくなる。このため、時間の経過とともに空間全体に広がる新規採択者を類似関係 (距離) でセグメント化する事は困難になると予想される。このアイデアは、市場に浸透した製品にターゲティング可能なロイヤルな顧客が形成されるという知見と整合しない。このため、新規採択者の類似関係の推移には、普及モデルや口コミの直接コミュニケーションのメカニズムとは異なるプロセスがあると期待できる。本研究の消費者の近さは、選好する製品カテゴリーの類似であり、普及モデルとは異なる概念を扱っているが、消費者の近さ

が売上推移パターンと関連する点は同じなので、消費者の分布の形状と売上推移パターンとの関連を見出せると期待する。

5.2.2 研究の課題

先行研究のレビューを踏まえて、第1の分析の課題としては、消費者の選好構造の空間的表現を示し、表5-1に掲げた性質の確認を行う。まず、表5-1の(B1)について、一般に、類似関係は距離の三角不等式を満たさないため、類似関係が導くプロダクトマップは、そのまま距離空間上に描画できないという課題がある。このため、類似係数から布置を作成する方法として、多次元尺度構成法などのモデルを用いることが多い。第3章と第4章において、行動面から製品カテゴリーの類似関係を評価するのに適していて、また製品選択を有意に説明する交互作用統計量は、2項分布の相関係数であるため（広津, 1992）自然な線形空間への対応が存在する類似係数である。表5-1の(B2)については、第3章でクラスター分析の結果として、製品と消費者の分類が対応する事をカイ二乗等計量により確認している。次節において、その理由について考察する。表5-1の(B3)については、交互作用統計量が評価する類似関係は、消費者 - 消費者、または製品 - 製品の距離であり、消費者 - 製品の距離に対応しない。また、第3章でみた類似係数として距離空間に自然に対応するが製品カテゴリーの類似関係を推定するのに適さない類似係数もあった（simple matching 係数）。このため、(B3)は実証において確認すべき事項と考える。

第2の分析における新製品の新規採択者に関する応用事例の分析では、普及理論における浸透あるいは接触の概念が援用できるどうかに着目する。製品と消費者の距離が製品選択の強さを示すのであれば、新規採択者の配置の推移は、潜在顧客への浸透の仕方を表していよう。このため、新規採択者の配置の推移を見れば、普及モデルと同様に売上推移パターンに関連づける知見を見出せよう。

5.3 分析手法

5.3.1 消費者の選好構造の空間的表現の構成

第3章と第4章の結果を踏まえ、消費者セグメントの類似評価にも、製品カテゴリーの類似関係の評価と同様に、交互作用統計量を用いる。行動データから類似評価を行って空間的

表現を行う手法としては、カイ二乗距離を用いる対応分析などがあるが、類似評価の手法としての交互作用統計量の優位性は第3章で確認した通り。また、交互作用統計量は以下に示す通り空間的表現が可能な類似係数である。消費者間の交互作用統計量の計算法は、製品間の交互作用統計量の計算過程における消費者と製品を入れ換えた計算とする。

まず、消費者間の交互作用統計量による類似評価が、2項分布の相関係数であり、集計対象の製品に軸が対応する直交空間に自然に配置できる事を示す。

消費者 n 人の m 個の製品の購買履歴データを表す行列 $M = \{m_{i,j} | m_{i,j} \in \mathbf{R}, i = 1, \dots, n; j = 1, \dots, m\}$ について、行和を $u_i = \sum_k m_{i,k}$ とする。また、 $p_i = \frac{u_i}{m}$ 、 M の第 i 行を \mathbf{m}_i とすると、消費者間 $i1$ と $i2$ の間の交互作用統計量 $s_{i1,i2}$ は下式 (5.1) の通りになる (広津, 1992)。

$$s_{i1,i2} = \frac{\frac{1}{m}(\mathbf{m}_{i1}, \mathbf{m}_{i2}) - p_{i1}p_{i2}}{\sqrt{p_{i1}(1-p_{i1})p_{i2}(1-p_{i2})}} \quad (5.1)$$

ただし、 $(\mathbf{m}_{i1}, \mathbf{m}_{i2})$ はベクトル $\mathbf{m}_{i1}, \mathbf{m}_{i2}$ の内積を表す。式 (5.1) の右辺の分子は、平均が p_{i1} と p_{i2} の2項分布に従う2種類の変数の積の平均から期待値の積を差し引いており、共分散の計算となっている。分母は同じ2変量の分散の積の平方根である。よって、消費者間 $i1$ と $i2$ の間の交互作用統計量 $s_{i1,i2}$ は、平均 $p_i = \frac{\sum_k m_{i,k}}{m}$, ($i = i1, i2$) の2項分布の相関係数に一致する。

ここで、消費者 i の m 次ベクトル表現 \mathbf{c}_i を下式 (5.2) で定義する。

$$\mathbf{c}_i = \frac{1}{\sqrt{mp_i(1-p_i)}} \mathbf{m}_i - \sqrt{\frac{p_i}{m(1-p_i)}} \mathbf{e} \quad (5.2)$$

ただし、 $\mathbf{e} = (1, \dots, 1)$ の m 次行ベクトル。

このとき、内積の計算により容易に下式 (5.3) の成立を確認できる。

$$s_{i1,i2} = (\mathbf{c}_{i1}, \mathbf{c}_{i2}) \quad (5.3)$$

式 (5.2) は、 m 個の製品が各軸に対応する直交座標系のベクトルとして、消費者を配置した事を意味している。各座標の意味合は、座標値として製品の購入の有無によりゼロか1を対応させた m_i を設定しておき、式 (5.2) により基準化して単位ベクトルに変換している。単位ベクトル化したのは、相関係数である $s_{i1,i2}$ が $[-1, 1]$ に基準化されているスケールに合わせるためであり、消費者による購入量の多少の主効果の情報とは独立な類似関係のみを表現するためである。もちろん、消費者による購入量の多少に関心がある場合は、敢えて単位ベクトル化する必要は無く、ベクトルの長さで主効果、ベクトル間の角度で類似関係を表現でき

る。式 (5.3) は、消費者間の交互作用統計量による類似評価が内積に一致し、理想ベクトルポイントの意味で類似の配置を表す空間となっている事を示している。単位ベクトルで表した消費者 $i1$ と $i2$ の距離 $d(i1, i2)$ と内積 $s_{i1, i2}$ または (c_{i1}, c_{i2}) との関係は、内積空間の一般的な性質として次式で与えられる。

$$d(i1, i2)^2 = 2(1 - (c_{i1}, c_{i2})) \quad (5.4)$$

以上が、課題 (B1) の、製品カテゴリーと直交座標系が対応する空間への消費者の配置の記述である。

5.3.2 消費者セグメントと製品カテゴリーの対応について

この空間の構成上の課題は、各軸に製品を対応させる際に、製品カテゴリーの類似関係を考慮して軸を斜交回転するといったステップを組み込まなかった事である。それにも関わらず、第3章で見たとおり製品カテゴリーの分類（上記空間の軸構成）と消費者セグメントの分類（上記空間の座標）が対応するのは、軸を斜交する効果より製品の購買のゼロか1かという差異の効果が大きいと解釈できる（ただし、ゼロか1は式 (5.2) でスケール調整されている）。例えば、 $\cos(80^\circ) = 0.17$ から製品間の交互作用統計量が0.17のときに該当する製品の軸を直交から 10° 傾けるというアイデアが浮かぶ。しかし、斜交の効果は製品の購買の有無よりも影響が小さいと思われ、製品に対応する軸を直交座標系としても、意味のある分析が導いているものとする。ところで、軸の斜交回転のプロセスを組み込むには、技術的な難しさがある。すなわち、製品の軸の斜交の調整を行うと、対応する消費者の類似関係にも影響する。このため、消費者間の交互作用統計量を計算する際には、製品間の類似は考慮しないという交互作用統計量の計算式自体を見直し、製品と消費者の類似評価の修正を収束するまで繰り返す必要がある。そこで、軸の斜交については今後の課題とする。以上が、第3章で確認した消費者セグメントと製品カテゴリーの対応に基づいて、消費者セグメントの空間的表現についての考察である。

5.3.3 消費者の配置と選好の強さの検証法

消費者の選好構造の空間的表現が備えるべき性質、(B3) 消費者の座標と選好の強さの対応について、特定の製品毎に検証を行う。

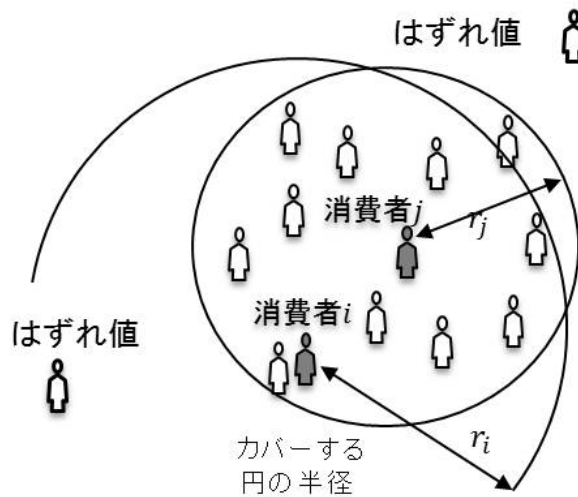


図 5-2. 中心的採択者 概念図

まず、概念図 5-2 の通り、分析期間の購買履歴において、特定の製品を購入した消費者をピックアップし、お互いの位置関係を交互作用統計量（式 (5.3)）あるいは距離（式 (5.2)）として評価する。次に、これらの消費者の分布の中心的採択者を N 人選択する。選択する基準は、ピックアップした消費者の中心に位置した消費者ほど、これらの消費者をカバーする半径が小さいという関係を利用して、中心的採択者の候補の消費者 i から消費者の $p\%$ をはずれ値として除いた、残りの消費者をカバーするための半径 r_i が最も小さい消費者の N 人とする。全採択者をカバーしないのは、はずれ値となる小数の採択者に影響されない意図がある。また中心的採択者を複数名とするのは、偶然選ばれた中心的採択者によって結果が左右される事を避けて、安定した分析を得る狙いがある。パラメータとなる N 人と $p\%$ の値は、実証において妥当な値を評価して決定する。

次に、概念図 5-3 の通り、特定の製品の中心的採択者の配置の重心からの距離（あるいは内積による類似評価）が、製品の選好の強さを表している事を確認するために、中心的採択者から同心円状に、布置する消費者数と、このうち該当する製品を採択した消費者数を集計し、同心円の半径をヨコ軸とした度数分布を描画する。予想として、同心円が広がるにつれて円に含まれる消費者は二乗オーダーで増えるが、

$$\text{採択率} = \text{採択者数} / \text{布置する消費者}$$

は低下すると考える。これを確認する。

中心的採択者を用いない別の手法として、該当する製品に対応する軸との距離を利用す

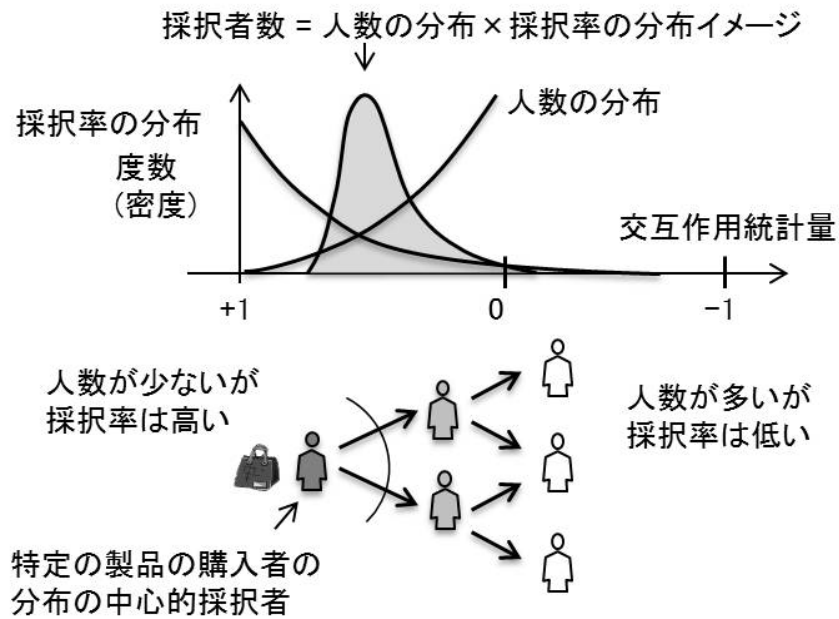


図 5-3. 製品（中心的採択者）と消費者の配置と製品選択確率の概念図

るアイデアが考えられる。しかし、第 4 章の研究結果より、消費者の将来の選択を予想するには、複数の製品間の類似関係を利用する必要がある事が分かった。第 5 章で提案している空間は、製品に対応する軸が直交したままであり、製品の類似関係の情報が失われている。例えば、製品 a と b の 2 種類あり、製品 a と b が交互作用統計量による評価で類似していれば、製品 a を購入しやすい消費者の中心は、製品 a の軸から製品 b の軸へ少し傾いた位置にあると思われる。製品 a を既に購入した消費者、あるいはこれから購入する消費者の中での選好の強さを比較する場合には、複数の製品間の類似関係を利用することが重要であると考え、中心的採択者の配置を直接求めて利用する手法を用いる。将来的には、製品の類似関係に応じて軸を斜交する課題に対応し、中心的採択者を用いない評価を可能としたい。

5.3.4 新製品の新規採択者の配置の推移と売上推移パターン

新製品の新規採択者の配置の推移と売上推移パターンには、2 通りの見通しが考えられる（図 5-4）。まず、ケース 1 として、普及理論や口コミのように、近い消費者から波及的に製品が浸透していく状況が考えられる。直接コミュニケーションにおける接触の頻度が提案する空間表現の距離に対応していれば、新規採択者は当初の中心的採択者から同心円に沿って次第に遠ざかり、お互いの類似関係が薄れながら市場全体に拡散していくと想定される。別のケース 2 として、初期の採択者はバラエティーシーカーのように、お互いに強い類似関

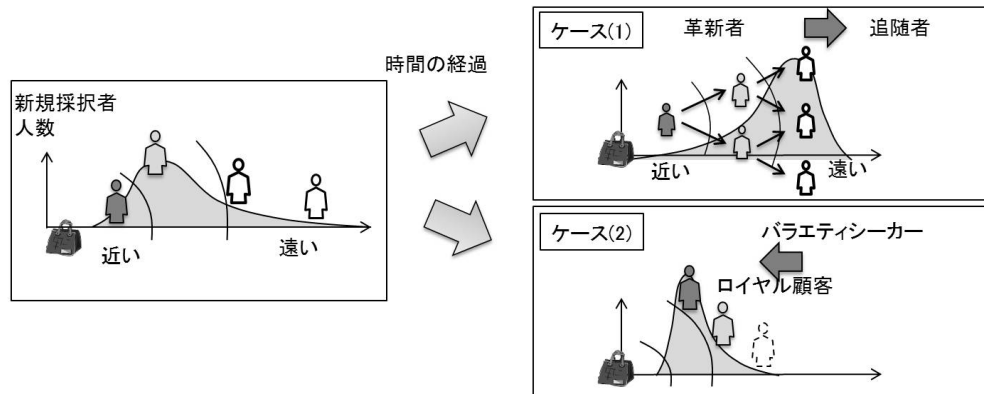


図 5-4. 想定する新規採択者の分布の推移パターン

係を持たない市場全体の製品に関心を示す消費者であるが、次第に製品の評価が定めるにつれて、新規採択者をカバーする同心円は小さくなり、やがて類似関係がお互いに強い同質的な消費者セグメントに限定されていくと想定する。

新規採択者の分布の推移を確認すれば、いずれのケースとなるか判断できる。まず、全期間のデータを用いて新規採択者の配置を決定する。次に、評価期間を期ずらししながら、該当する期間の新規採択者の交互作用統計量の分布を求める。 n 人の新規採択者がいれば、 $n(n-1)$ 個の交互作用統計量の分布を求め、その分布の四分位点の推移を見ることで、分布の形状が広がっているか、あるいは縮小しているかを評価する。

5.4 実証分析

実証分析では、第 1 の分析として、提案する空間的表現における消費者の配置が製品への選好の強さを表している事を評価する。第 2 の分析として、提案する空間的表現が実務的に有用である証左として、消費者の配置の分布と新製品の売上パターンとの関連を評価し、新製品のプロモーション戦略に有益な知見を新たに導く。

5.4.1 データ

2 つの分析では、第 3 章と第 4 章で用いたデータと同様に、大手セル・レンタル CD チェーン店から個人が特定できないように加工した上で提供頂いた、ID 付き購買 POS データを用いる。データ数は、音楽 CD 販売の購買履歴 707,085 件 (消費者 101,307 人、製品 37,297 個)。購買の期間は 2002 年 11 月 1 日から 2003 年 12 月 21 日 (13 ヶ月)。データを収集し

た店舗は、東京と大阪エリアの計 5 店である。市場の特徴として、値引きなどのプロモーション効果の影響が少なく、繰り返し購買がほとんど無いという特徴がある。また、衝動買いが少なく、各消費者の好みが主な購買要因とされる（一般社団法人日本レコード協会, 2008,2009）。

評価事例は新製品と既存製品の売上の多い 50 製品に限定する。音楽 CD のタイトル毎の売上は、上位から指数的に減少するため、売上げ上位の製品に重みを置いて評価する事が重要と考える。また新製品の定義として、購買履歴の最初の 3 ヶ月間に購買が全く無い製品を新製品として扱う。

5.4.2 分析結果 1：消費者の配置と製品への選好の強さの関係評価

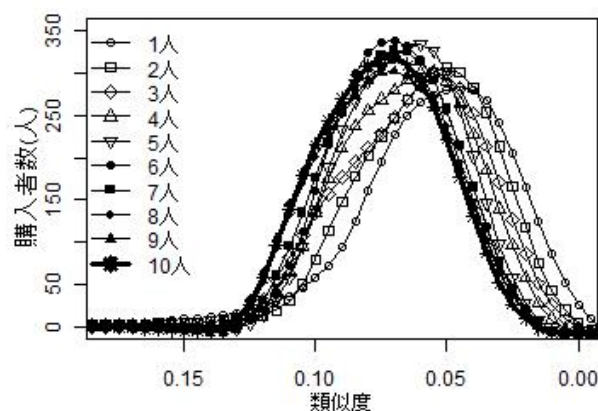


図 5-5. 中心的採択者数パラメータ評価

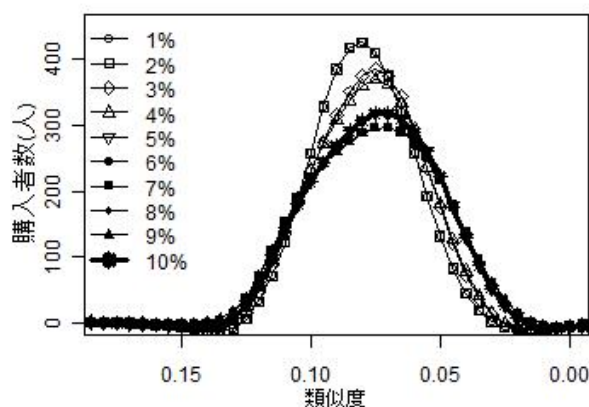


図 5-6. はずれ値人数比率パラメータ評価

各製品の中心的採択者を決めるため、「5.3.3. 消費者の配置と選好の強さの検証法」の図 5-2. で記した、中心的採択者の人数 N 人と、はずれ値の割合 $p\%$ の適当な水準を設定する。図 5-5 は、新製品の中で最も売上の多い「t . A . T . u .」について、人数 N を 1 人から 10 人と増やし、また、はずれ値の割合 $p = 10\%$ と固定して中心的採択者を求めた度数分布である。なお、分布の形状はカーネル密度推定法で平滑化している。それらの中心的採択者 k と他の「t . A . T . u .」購入者 i との交互作用統計量による類似評価 $s_{k,i}$ を、中心的採択者で平均 ($\frac{1}{n} \sum_k s_{k,i}$) して評価した交互作用統計量の分布である。先に、図 5-3 で考察した通り、購入者の交互作用統計量の分布は釣鐘型になるとの予想通りの結果を示したが、中心的採択者の人数を 1 人、2 人と増やすにつれて分布の形状の最頻値は左（交互作用統計量が高くなる）方向に推移し、最頻値の値も高くなり分散が縮小する好ましい傾向を示した。

そして、中心的採択者の人数を 10 人程度に置くと、分布の形状が安定した。他の製品についても同様の試行を繰返し、中心的採択者の人数を 10 人にとると安定した分布が得られる事を確認した。次に図 5-6 の通り、中心的採択者の人数を 10 人に固定し、今度ははずれ値の割合を 1% から 10% に変化させて購入者の分布を作成した。はずれ値の割合が増えると、最頻値の値は小さくなったが、10% のはずれ値を設けると分布の形状は安定した。同様に、他の製品についても試行を繰返し、中心的採択者を求めるパラメータとして、人数 $N = 10$ 人、はずれ値 $p = 10\%$ とした。

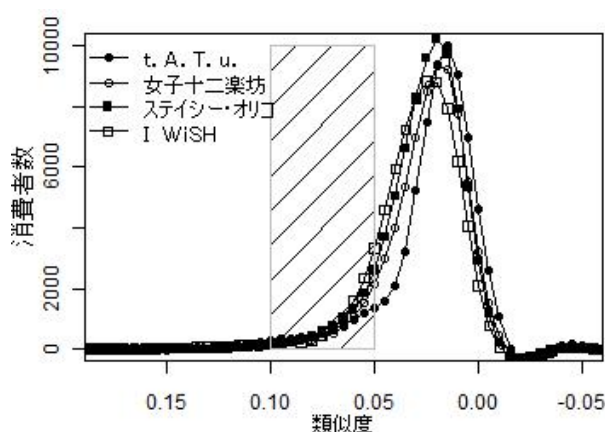


図 5-7. 潜在購入者の分布

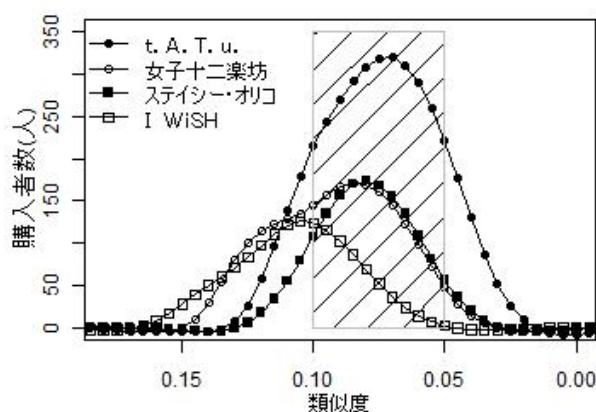


図 5-8. 購入者の分布

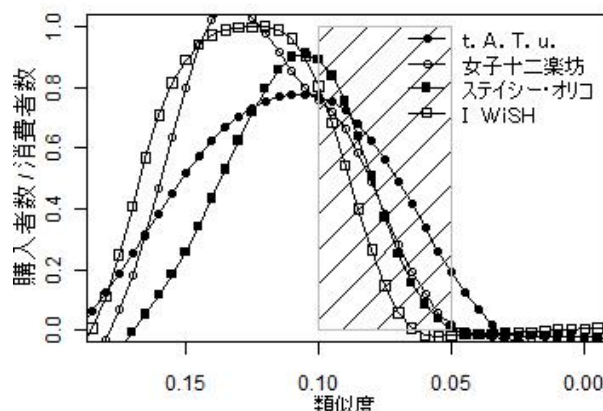


図 5-9. 採択率 = 購入者 / 潜在購入者の分布

次に、各製品について求めた中心的採択者からの平均的な距離を交互作用統計量で評価し、中心的採択者から同心円状に潜在顧客となる消費者の人数がどのように分布しているか評価した。新製品の売上 50 製品について分布の形状を確認したところ、概ね同じ形状となったので、例として新製品の売上上位の「t. A. T. u.」、「女子十二楽坊」、「スティシー・オリコ」、「I W i S H」について、評価した分布を図 5-7 に示している。より中心

に近い図の左の交互作用統計量の値 0.15 近辺では同心円の大きさも小さく、該当する人数もわずかであるが、網掛けした値 0.10 ~ 0.05 の区間で人数が増えはじめ、値 0.05 ~ 0.00 の区間で最頻値を示した。次に、図 5-8 に、該当する製品を購入した消費者に限定した度数分布を示した。今度は、図 5-7 で網掛けした区間と同じ値 0.10 ~ 0.05 の区間で最頻値を示した。最頻値が同心円の中心にずれたのは、より中心に近いと採択率も高いと思われ、購入者（図 5-8）/潜在購入者（図 5-7）で採択率を計算した結果をスプライン関数で平滑化して図 5-9 に示した。網掛けした区間より左では、分母となる人数が少なく分布の形状に安定性が無いと思われるが、網掛けした区間（交互作用統計量 0.10 ~ 0.05）では、同心円の中心から外に広がるにつれて採択率が単調に低下し、網掛けした区間をはずれると、採択率はゼロに近づいた。従って、これらの製品の有望な潜在顧客は、交互作用統計量による類似評価で値 0.05 の同心円内にいる消費者と考えられる。

表 5-2. アーティスト別の新規採択者の分布の特徴

アーティスト	t.A.T.u., スティシーオリコ	女子十二楽坊, I WiSH
購入者の類似評価	小さい, 異質的	大きい, 同質的
潜在購入者の分布	広い	狭い

アーティスト別の分布の形状の差異としては、図 5-8 において、「t . A . T . u .」と「スティシー・オリコ」の分布は、相互の類似が小さく、類似で評価した大きな同心円（図 5-8 で右方向）に分布しており、市場の広範囲に購入者が分布している事が分かる。このように、製品によって購入者の分布の広がり方に差異が見られる。

5.4.3 分析結果 2：消費者の配置の分布と新製品の売上パターンとの関連を評価

製品の売上は、競争空間において製品が位置する周囲の消費者の密集の程度に影響されると考えられる。市場において、そのような競合の影響が働いている事を確認するために、既存製品と新製品の双方からの影響の非対称性を確認する分析を行った。時系列的な因果として、既存製品は新製品に影響を与えるが、その逆は小さい。そこで、双方の影響の非対称性

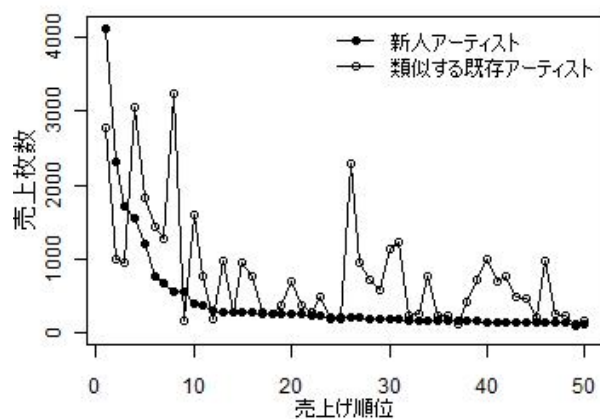


図 5-10. 新人アーティストの順位別
売上枚数と最も類似する
既存アーティストの売上枚数

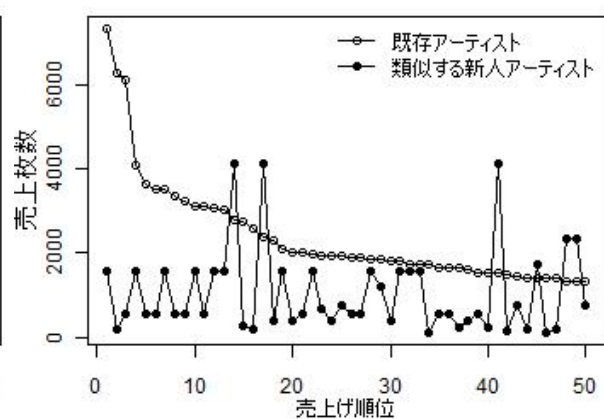


図 5-11. 既存アーティストの順位別
売上枚数と最も類似する
新人アーティストの売上枚数

が確認できれば、提示する競争空間において競合関係を評価する一つの根拠となる。まず、新製品の売上上位 50 位までのそれぞれの製品に対して、最も交互作用統計量が高い既存製品を求め、それぞれの売上を図 5-10 に示した。ヨコ軸は、既存製品内での売上順位である。新製品の売上 10 位以上において、対応する既存製品の売上が高くなっていて、相関関係が確認できた。同じ評価を新製品と既存製品を入れ替えて行ったところ、図 5-11 の通り、相関関係を確認できず、非対称性を示す結果となった。このため、新製品は周囲の類似関係の強い既存製品の影響を受けていると考えられる。最も直接的な影響は、既存製品の周囲に広がる顧客が新製品の新たな顧客になる事であろう。

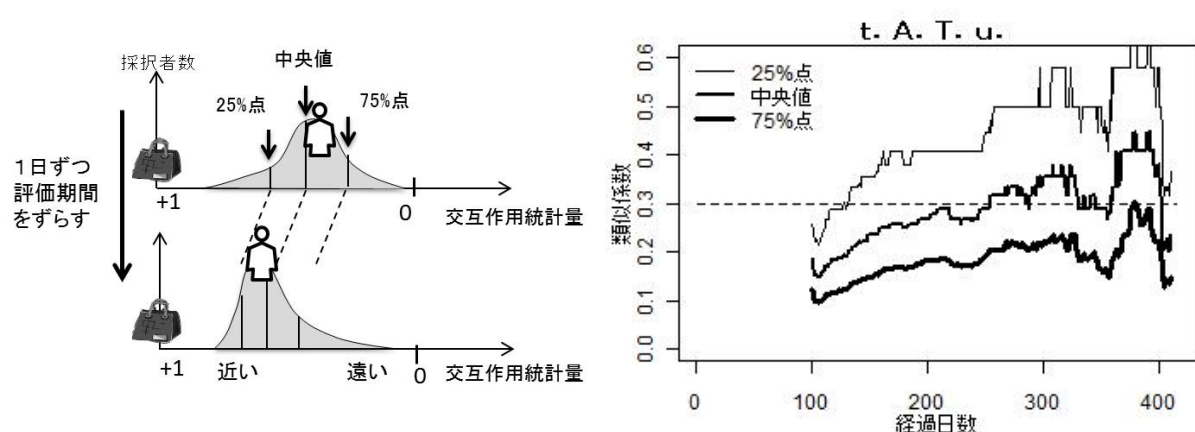


図 5-12. 新製品の採択者間の交互作用統計量の分布の推移

第 5 章の 2 つ目の課題である、消費者の分布の形状と売上パターンの関係を確認する分析を行った。ここで分布の形状とは、該当する製品の中心的消費者からの同心円上に位置する消費者の度数分布の形状を意味する。まず、データの全期間を使って消費者間の空間的配置

を設定して固定した。このため、時間推移によって消費者は空間を移動しないが、新製品の市場への浸透に伴い、新規採択者が出現する配置に変化があると期待する。例えば、住民の住所に変更は無いが、新規採択者の分布が全国へ広がるか、あるいは特定の地域の住民に集中するかを評価する。図 5-12 の通り、時間推移は 30 日間のウィンドウで測定し、1 日ずつずらしながら分布の四分位点の推移を求め、四分位点の値の変化により、分布の形状の推移を評価する。

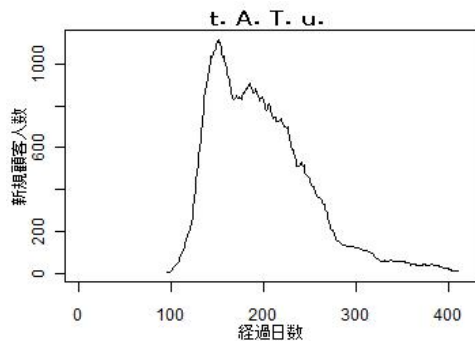


図 5-13-a. 新製品の新規顧客数推移
t.A.T.u. のケース

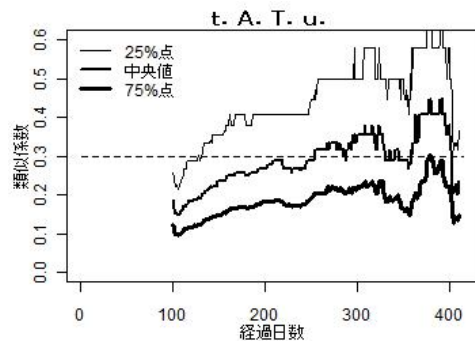


図 5-14-a. 新製品の顧客同質性推移
t.A.T.u. のケース

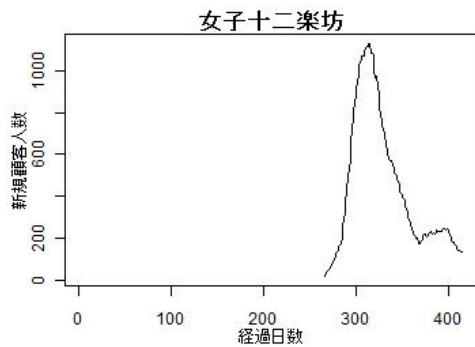


図 5-13-b. 新製品の新規顧客数推移
女子十二楽坊のケース

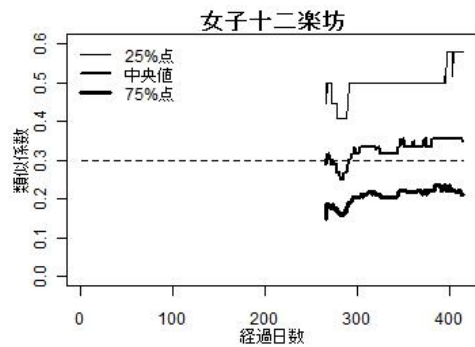


図 5-14-b. 新製品の顧客同質性推移
女子十二楽坊のケース

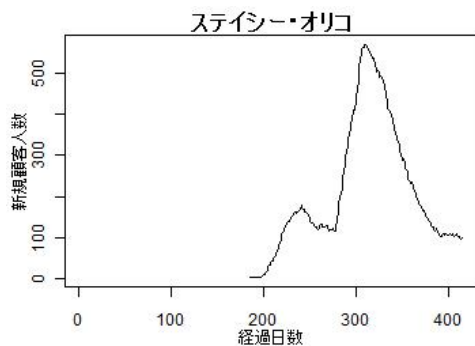


図 5-13-c. 新製品の新規顧客数推移
ステイシー・オリコのケース

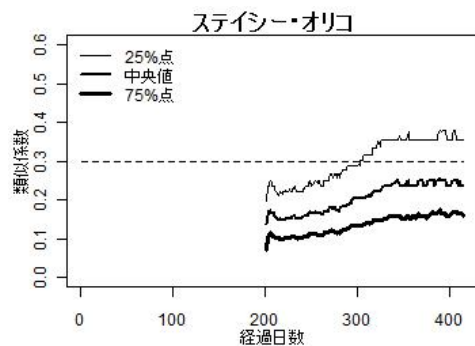


図 5-14-c. 新製品の顧客同質性推移
ステイシー・オリコのケース

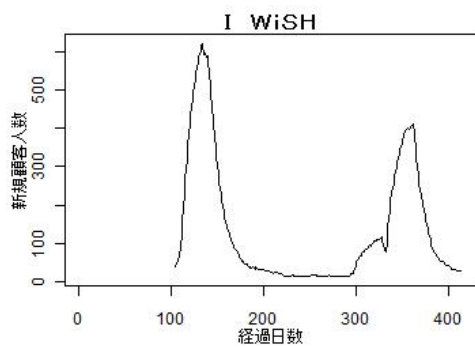


図 5-13-d. 新製品の新規顧客数推移
I WiSH のケース

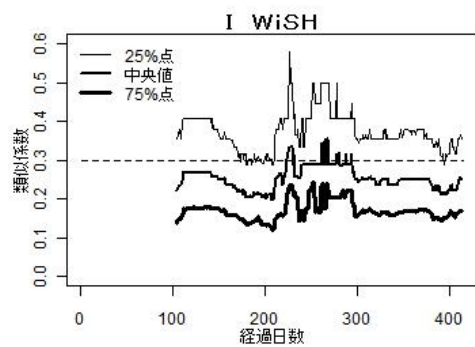


図 5-14-d. 新製品の顧客同質性推移
I WiSH のケース

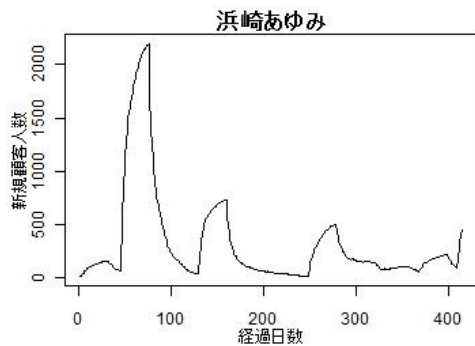


図 5-15-a. 既存製品の新規顧客数推移
浜崎あゆみのケース

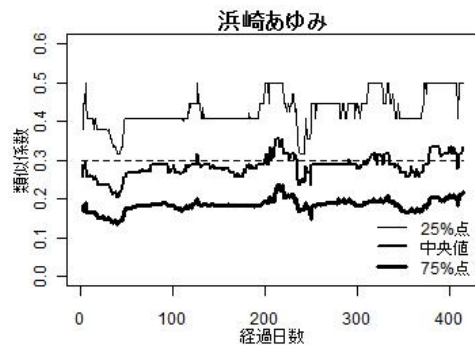


図 5-16-a. 既存製品の顧客同質性推移
浜崎あゆみのケース

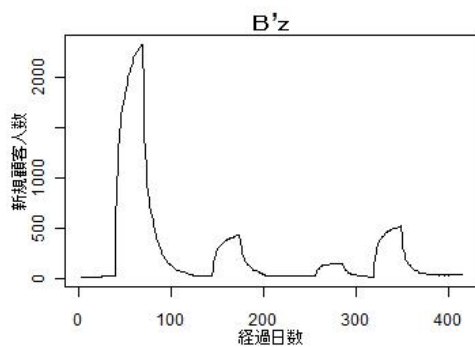


図 5-15-b. 既存製品の新規顧客数推移
B'z のケース

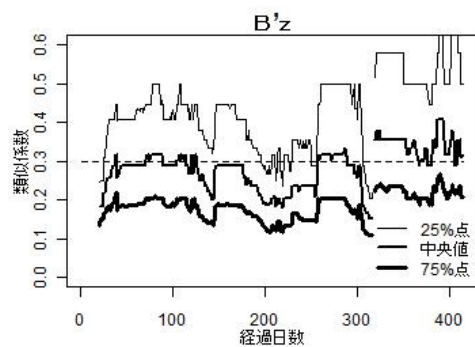


図 5-16-b. 既存製品の顧客同質性推移
B'z のケース

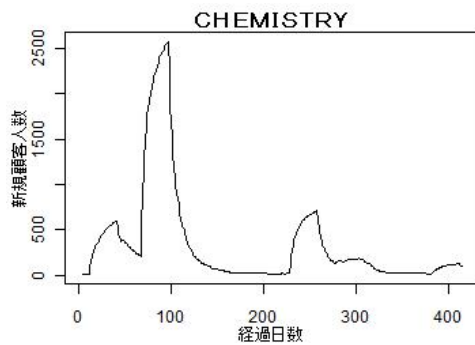


図 5-15-c. 既存製品の新規顧客数推移
CHEMISTRY のケース

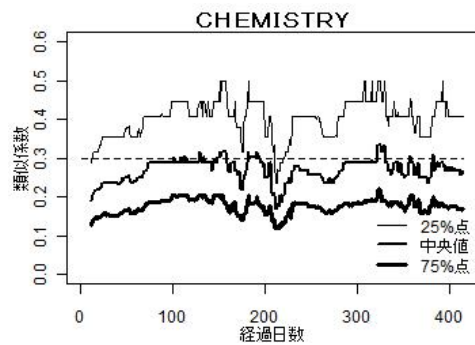


図 5-16-c. 既存製品の顧客同質性推移
CHEMISTRY のケース

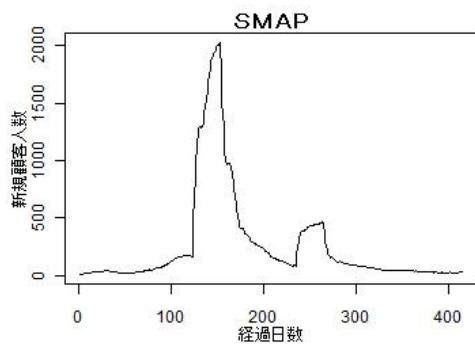


図 5-15-d. 既存製品の新規顧客数推移
SMAP のケース

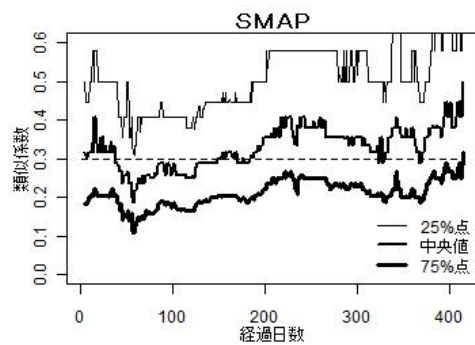


図 5-16-d. 既存製品の顧客同質性推移
SMAP のケース

図 5-13-a ~ d は、新製品上位 4 製品の P O S データにおける経過日数毎の売上推移パターンを示している。「t . A . T . u .」と「ステイシー・オリコ」が比較的ゆっくりと売上を伸ばして減少していくのに対して、「女子十二楽坊」と「I W i S H」は短期で売上を伸ばして減少していく。「I W i S H」に売上のピークが 2 箇所あるのは、異なる新譜の売上がある事を示している。対応する製品の、採択者の分布の四分位点の推移を図 5-14-a ~ d に示している。「t . A . T . u .」と「ステイシー・オリコ」は比較的ゆっくりと消費者間の類似が強くなり、新規採択者の四分位をカバーする同心円の半径は小さくなり中心に近づいている。これに対して「女子十二楽坊」と「I W i S H」は、当初から消費者間の類似は強く、同心円の半径も小さい。なお、「I W i S H」の図 5-14-d. において変動が大きいのは、売上が少ない期間で推計しているためである。

既存製品についても同様に、売上パターンの推移を図 5-15-a ~ d に、採択者間の交互作用統計量の分布の四分位点の推移を図 5-16-a ~ d に示した。新製品はやがて既存製品と分類されるので、既存製品の特徴は、新製品が推移した結果の到達点と解釈できる。図示した既存製品は、いずれも「I W i S H」のように売上が短期に集中してノコギリの歯の形状を示し、類似評価の四分位点の水準も高くなった。

5.5 結果と応用および今後の課題

5.5.1 結果

分析 1 において、図 5-8 の通り、構成した消費者セグメントの空間的表現において、消費者と製品の中心的採択者との類似を交互作用統計量の値を横軸として採択者の度数分布を描くと、製品から一定の距離だけ離れた所で採択者数が最も大きくなる釣鐘形の分布をしていることを確認できた。これは、消費者が製品に近づく（交互作用統計量の値が大きくなる）につれて、選択確率が高くなる要因と（図 5-8）と、消費者が製品から遠ざかるにつれて、大きくなる同心円に含まれる消費者数が増える要因（図 5-9）の効果と解釈できる。また、評価対称となる製品を変更しても、安定して同じ形状を確認できた。以上の結果より、提案した空間的表現に期待する性質 (B3) 消費者の座標と選好の強さが対応する、を確認できた。既に確認した、(B1) 製品カテゴリーに直交座標系が対応する、(B2) 消費者セグメントと製品カテゴリーが対応する、と併せて、消費者セグメントの選好を表現する空間として

の妥当性を確認できた。

分析 2 において、新製品の売上と、その新製品と最も類似する既存製品の売上げには、プラスの順位相関が見られた（図 5-10）。ただし、類似の評価には交互作用統計量を用いている。この結果より、新製品と補完・代替関係にある既存製品の存在が、新製品の売上げにプラスの効果を及ぼすと解釈できる。また、逆に新製品と既存製品を入れ替えて同じ分析をした結果、有意な順位相関が見られなかった（図 5-11）。この結果の解釈として、新製品は周囲に多くの顧客を持つ既存製品の近くに位置することで、それらの既存製品の顧客を取り込みやすい地点があると考察できる。また、この結果は成功している既存製品に類似するポジションが、売上にプラスに働く 2 匹目のドジョウ戦略が有効である事を示唆している。

5.5.2 プロモーション戦略への応用

表 5-3. 2 つの典型的なプロモーション戦略

	独自市場開拓	2 匹目のドジョウ・マス広告戦略
例	t.A.T.u., ステイシーオリコ	I WiSH, 女子十二楽坊
売上推移	長いライフサイクル	短期ライフサイクル
顧客の分布	広範囲・多数	狭い・ロイヤル顧客
要件	製品本来の魅力	効果的広告

新製品の中で、「t.A.T.u.」, 「ステイシーオリコ」といった新製品は、市場に浸透するにつれて新規に採択する消費者の間の交互作用統計量の値が大きくなり、すなわち同じ選好を持つ消費者の比率が増える同質化の傾向を確認できた（図 5-14）。この結果は、初期の採択者から同心円状に新規採択者が波及するという普及理論や口コミの知見と真逆であるが、交互作用統計量は直接コミュニケーションの隣接関係とは別の概念を捉えているため、普及理論の知見に矛盾してない。解釈としては、当初はバラエティー・シーカーなど様々な選好の消費者に注目されるが、市場における評価が定まると、次第に同じ選好を持つロイヤル層が形成されると考えられる。一方で、「女子十二楽坊」と「I WiSH」いった新製品は、当初より消費者の同質化が見られ、既存製品と同じような短期の売上パターンを示す製

品も見られた。この傾向は、既存製品で顕著となり、新譜の売り上げは短期期間に集中して急激に減少するノコギリ型となっている（図 5-15）。

「女子十二楽坊」と「I W i S H」は、当初よりテレビへの露出が多いアーティストであった。洋楽に分類され活動の拠点が日本では無い「t.A.T.u.」や「ステイシーオリコ」と異なり、「女子十二楽坊」と「I W i S H」はTV番組とのタイアップや歌謡番組への出演で、新譜の発売時から消費者の認知が高かったと考えられる。表 5-3 に、これらの2 類型の特徴をまとめた。「t.A.T.u.」や「ステイシーオリコ」のように、製品本来の価値が次第に認められ市場全体から広く顧客を獲得するのは望ましいが、そのような製品を開発する困難がある。一方で、「女子十二楽坊」と「I W i S H」のようにマス広告を用いて顧客を開拓すると、広告費が高く、また製品のライフサイクルが短くなり易いという課題がある。「女子十二楽坊」のプローションに多額の投資を行った「ミューチャー・コミュニケーションズ」（東京都港区）はその後に自己破産申請し、社長はインタビューに対して「女子十二楽坊」への投資が過大であったと語っている。

本手法は、これらの新製品の売上パターンをモニタリングし、効果的なプロモーションを行うのに役立つ。例えば、「t.A.T.u.」の初期のプロモーションでは、顧客をバラエティーシーカーにターゲティングし、音楽CDに關与の高い消費者が接する雑誌やFMラジオやレコード店でのプロモーションを行い、一定期間後からは、顧客ターゲティングを変更してマス広告を行うといった戦略が考えられる。また、「I W i S H」のように市場への浸透を急速に行うのであれば、「I W i S H」を認知している消費者を初期からターゲティングする戦略が有効と思われる。本手法を用いれば、初期の購入予約の履歴から、比較的容易に顧客ターゲティングできよう。このように、本手法は新製品の売上パターンの類型化を捉え、販売状況をモニタリングしながら顧客ターゲティングを行うために応用できると考える。

5.5.3 学術的貢献と今後の課題

第5章の研究を通して、消費者の製品間の類似性に関する評価は、企業の製品ポジショニング戦略や、あるいは、広告やインストアプロモーションなどのマーケティングコミュニケーションを用いて戦略的に認知させたり、変化させたりすることができる、統制可能な評価要因である事を示唆している。本研究の成果として、以上の通りマーケティングにおける製品カテゴリーの類似関係の役割を見出せた。また、提案した消費者の選好構造の空間的表

現を用いる事で、時間的に連続した市場の変化を観察することができた。この空間的表現は、市場を単に観察するのに適しているだけでなく、線形空間を前提とする様々な統計手法が利用できるという点において、拡張性の高い分析フレームになろう。

今後の課題として、結果が実証データに依存している限界があるため、様々な市場において同様の結果が得られる事を検証することが重要と考える。また、空間的表現においては、消費者の類似を評価する際には製品の類似を考慮しないという、交互作用統計量の本質的な課題への対応が重要であると分かった。分析結果は、幅広い顧客を獲得する好機は、新製品の市場投入時である事を示唆しており、顧客の同質化が短期間に進む製品と、時間をかけて市場に広く浸透する2つのタイプの製品が見られた。ライフサイクルを長期化させる観点から、新製品の市場投入時のプロモーション戦略の類型と、その選択基準を明らかにする事を今後の課題としたい。

第6章 結論

6.1 結論

第1章では、多様な消費者の選好構造を捉えるために、消費者のカテゴリーベース処理に着目し、消費者の選好構造を空間的表現として捉えるために次の研究目的を設定した。第1の研究目的は、行動面から消費者の製品カテゴリーの類似関係を推定する手法の妥当性を示す事。第2の研究目的は、製品カテゴリーの類似関係が製品選択に与える影響を明らかにする事。第3の研究目的は、製品カテゴリーの類似関係と消費者セグメントの対応を用いて消費者の選好構造を空間的に表現し、また消費者の選好構造と売上パターンとの関係を明らかにして新製品のプロモーション戦略への応用を提案する事。以上の3つの研究目的を設定した。

第2章では、第1章で設定した3つの研究目的に関連する先行研究として、ピースミール処理とカテゴリーベース処理の概念と特徴の対比、カテゴリー概念とカテゴリーの類似関係の定量的表現、競争市場構造分析による製品カテゴリーと消費者セグメントの対応手法と空間的表現、および消費者の異質性の扱い、製品カテゴリーの類似関係を用いた製品のクラスター分析手法と評価基準について、各研究分野のレビューを行い、それらの内容を踏まえて、(1) 消費者の行動面から製品カテゴリーの類似関係を推定する手法の提示、(2) 製品カテゴリーの類似関係と製品選択との関係を示すモデルの提示、(3) 競争市場構造分析の空間的表現として多様な消費者の選好構造を捉え、消費者の選好構造と売上との関係について新たな知見を示す、以上の3点を研究課題として設定した。

第3章では、多様な消費者の製品選択を捉える目的から、製品カテゴリーの類似関係を表現するための適切な類似係数を選択する研究を行った。これまでに多数の類似係数が提案されているが、類似係数の選択基準は示されていない。そこで、交互作用統計量を含む16種類の類似係数で分類を作成し、教師なし分類の妥当性評価手法であるクラスターバリデーションを用いて各分類結果の妥当性を比較した。クラスターバリデーションの評価基準には

再現性・均等性・外的基準・内的基準の定量的な基準と、定性的な解釈可能性を用いた。特に定量的評価においては分類の2-フォールド交差確認法を取り入れて評価の信頼性を高めた。音楽CDの個人ID付き販売POSデータを用いて実証分析を行った結果、交互作用統計量と相関が高い類似係数ほど妥当な分類を与える事を確認した。交互作用統計量が2項関係であるのに対し、主効果は各観測対象に固有の属性である。このため、観測対象に直線的な全順序を与えてしまい、分類というよりランキングの機能を果たしていると考えられる。主効果と相関の高い類似係数を用いて階層的クラスタリングを行うと、空間収縮的と呼ばれるチェーンの形状をしたデンドログラムを生成し、各分類における要素数が著しく不均等になった。これらの結果から、主効果と独立な交互作用統計量を用いて消費者のカテゴリー構造を捉えるのが適切である事を検証により示した。

第4章では、第3章の結果を踏まえて、交互作用統計量を用いて製品カテゴリーの類似関係を評価し、この類似関係が製品選択に与える影響を研究した。交互作用統計量は、購買経験を条件とする条件付き確率と、期間併売との関係を与える点に着目し、交互作用統計量を説明変数として、消費者の将来の購買の確率を推定する算式を導いた。消費者の過去の購買経験が将来の製品選択に影響を与えるという考え方は、過去の経験によって得た知識に基づいて消費者がカテゴリーベース処理を行って製品を理解し、その結果が製品選択に影響を与えていると解釈することもできる。音楽CDの購買履歴の実データと模擬データを用い、導出した購買確率を用いて将来の製品選択を予想したところ、推定が有意である事を検証できた。また、市場の競争構造としてレコードレーベルによる選択と集中戦略が観察される音楽CD市場では、同じレコードレーベル内のアーティストの類似関係が高く、第3章で提案した購買確率の推定式による売上の相乗効果が観察されるならば、同じレコードレーベル内であろうと推測し、これを確認する数値実験を行った。その結果、推測の通りの結果を得て、提案した購買確率の推定式の妥当性を示す内容であると解釈した。

第5章では、競争市場構造分析として、多様な消費者の選好を空間的に表現して捉える研究を行った。第3章と第4章の結果から、消費者の行動面から製品カテゴリーの類似関係を捉える評価式として交互作用統計量が適切であり、また逆に、交互作用統計量を用いた製品カテゴリーの類似関係評価が消費者の製品選択に関係する事が分かっている。そこで、さらに踏み込んで、交互作用統計量を用いて評価した製品カテゴリーと消費者セグメントの対応関係を空間的に表現し、競争市場構造分析の観点から消費者の選好構造を表現している事の確認を行った。交互作用統計量は2項分布の相関係数であるため、その値を内積とする直交

座標系を構成することができる。具体的には、各製品を各軸に対応させた直交座標系に消費者をベクトルとして配置し、消費者間の交互作用統計量による類似評価が内積に一致する理想ベクトルモデルとしての空間配置が、先行研究を踏まえて独自に設定した競争市場構造分析の空間的表現としての3つの評価基準を満たしている事を実証で確認した。応用として、新製品の新規採択者の提案空間における分布の推移を評価して、新規採択者同士の距離が次第に近くなり、購買パターンが同質化する傾向がある事を知見として見出した。また、この変化が、売上推移のライフサイクルの短期化の要因であるとの解釈を示し、新製品のプロモーションにおける応用を考察した。

第6章では、本研究のまとめとして研究目的と課題および成果を整理し、また学術貢献と実務的貢献、今後の課題について記した。

6.2 学術的貢献

行動面からの製品カテゴリーの類似関係評価

消費者の知覚から製品カテゴリーの類似関係を評価する Tversky(1977) の対比モデルと同様に、消費者の行動面からも製品カテゴリーの類似関係が評価できる事を示せた。購買行動を目的としない知覚情報からの製品の類似評価と、購買行動を目的とした場合の製品の類似評価は、必ずしも一致しない可能性はあるが、マーケティングにおける関心は購買行動にあり、この点において行動面からの製品カテゴリーの類似関係評価は、マーケティングの研究目的に沿うものである。

製品カテゴリーの類似関係評価における交互作用統計量の妥当性

製品カテゴリーの類似関係を捉える類似係数として、主効果は不要であり、交互作用統計量が適切である事を明らかにした。類似係数は、多次元尺度構成法やカーネル法などに利用されることから、主に統計学において類似係数の選択や評価基準に関する様々な研究があり (Janson & Vegelius, 1981; Baulieu, 1989; Duarte et al., 1999; Wadel & Kamakula, 2000; Albatineh et al. 2006)、これらの研究では、10個～20個前後の類似係数を網羅的に相対比較する研究が行われているが、類似係数の選択基準は提示されていなかった。

本研究では、消費者セグメントと製品カテゴリーを作成するのに用いる類似係数として、主効果がクラスタリングにおいて空間の歪みを生じる要因となり、交互作用統計量が優位な類似係数である事を、実証分析として示した。このため、例えばカイ二乗統計量を用いるコレスポンデンス分析を大規模データに適用すると、主効果の感応度が交互作用統計量を打ち消す傾向があり、分析結果の主効果の要因が強くなるといった解釈が可能となった。以上の考察は、類似係数を用いる分析において有用な知見と考える。

また、交互作用統計量は2項分布の相関係数であるため、交互作用統計量の値を内積とする直交座標系が存在し、この特徴を用いて理想ベクトルモデルとしての競争市場構造分析の空間的表現が自然に対応する。このため、消費者の選好構造を空間的に表現するための指標として交互作用統計量が優れた利点を持つことが分かった。

本研究では、消費者の内面にあって観測できない製品カテゴリーの類似関係と、その知識を用いた情報処理による製品選択に関心があり、提案する手法の是非を直接的に検証できない。このため、単純で明示的な設定を置いて、そこから導かれる結果の妥当性や不都合さを確認するという検証的な手続き以外に研究を進める仕組みは無いと考える。数理モデルは、その単純で明示的な設定を記述する手続きとなる。

交互作用統計量を用いた消費者の選好構造の空間的表現

本研究で提案する空間的表現では、各軸が製品に対応し、消費者が布置として示される、解釈の容易な構成となっている。また、線形空間として表現された消費者の選好構造は、様々な統計手法を適用し易く、市場構造の分析フレームとして利便性が高い。

提案する競争空間を用いた応用事例として、新製品の新規採択者の分布を時系列で評価し、新規採択者のお互いの類似係数の分布の範囲が狭くなるという知見を見出せた。口コミの伝播や普及理論の浸透の考え方に沿うと、最初に製品に近いポジションにいる消費者が選択し、次第に広く市場全体に及ぶと予想するのが自然であるが、本研究では、この逆の結果を得た。新製品の普及を分析する際に、この知見は重要であると考え。Bass (1969) の普及モデルでは、潜在顧客数を所与として全ての潜在顧客に製品が浸透するまでの推移を説明する。この際に、予め潜在顧客数を予測するは困難であるが、本研究の知見は、潜在顧客数が縮小する傾向にある事を示している。このような知見が見出せたのは、時系列の連続的変化を捉え易い空間的表現として、個々の消費者の選好構造を適切に表現できたため、個々の

消費者の購買の集計である売上との関係を示せたと考える。

本研究の主な目的は、多様な消費者の選好構造を捉える事にある。独自の選好を持つ消費者は、大多数の購買行動に倣うのではなく、自らの価値観に基づいて製品の評価を行うため、多くの製品知識を自ら集めて合理的に情報処理する必要があるが、消費者の情報探索や処理能力には限界がある。そこで、製品に対する既存の知識をカテゴリー化し、類推を行うことで認知的努力を小さくする情報処理をしていると予想し、製品カテゴリーの類似関係に対応する消費者の選好構造を空間的に表現した。前述の新たな知見の発見は、提示した空間的表現が、この目的に沿っていたとの解釈を与える。

6.3 実務的示唆

類似係数を用いた判別手法への応用

類似係数は、実務において良く利用される指標である。その背景として、データの制約から離散的な限られた情報を最大限に活用する必要があったり、コンピューターシステムに関するインフラの普及により、WEBサイトの閲覧履歴や購買履歴のような2値データの蓄積が増えている事がある。例えば、生命医学におけるゲノム解析 (Duarte. et al.,1999) や、コンピューターシステムとしてのレコメンドシステム (芳賀, 2008) や、マーケティングにおけるカテゴリー化 (Wadel & Kamakula, 2000) や、テキストマイニング (Shawe-Taylor & Cristianini, 2004) などの、様々な領域で類似係数が利用されている。

本研究で明らかにしたのは、消費者のカテゴリー処理を捉えるための類似係数の選択基準なので、ゲノム解析などの人間の判断が介在しない事象に対しては、また違った評価基準もある。とはいえ、類似係数の評価において、交互作用統計量と主効果の評価軸の有用性は、全ての実務的領域においても同様と考える。第3章で見たように、適切な類似係数を選択したかどうかで、分類結果は大きく異なる。本研究の成果により、これらの応用領域での類似係数の選択の指針を提示できたと考える。

レコメンドシステムへの応用

コンピューターシステムとしても、マーケティングツールとしても、レコメンドシステムは重要なツールとなっている。第4章における交互作用統計量を用いた消費者の製品選択の

推定確率の導出は、なぜ類似係数が製品選択を説明するかの解釈を与える。これまで、レコメンドシステムで用いる協調フィルタリングのアルゴリズムは、概念的に「類似する消費者の過去の購買を参照する」と説明したり、「相関係数の欠損値を推定するのと同様に」といった表現で語られていた。これに対して本研究では、交互作用統計量を購買履歴条件付き同時確率とみる事により、交互作用統計量で製品選択を説明する新たなアプローチを示す事ができた。

このアプローチに沿って考えると、消費者の過去の購買履歴が将来の製品選択を決定するという、逐次的な消費者の選好の変化として、協調フィルタリングを捉えることができる。また、推定式的前提条件や導出過程が明快であるため、推定式の適用結果の優劣の要因を容易に考察できる。このため、本研究成果を踏まえれば、レコメンドシステムの設計者は、経験的手法を用いる場合のような網羅的な試行錯誤に陥ることなく、段階的にレコメンドシステムを構築したり、改善することができよう。

類似関係を用いたプロモーションへの応用

第4章では、音楽CD市場の分析結果として、代替関係にあるアーティストの大半のペアは、同じレコード会社に所属している事を確認した。また、第5章では、代替関係にある製品の売上枚数の相関が高くなるという、いわゆる2匹目のドジョウ戦略の有効性を確認した。したがって、代替関係にある製品のそれぞれのプロモーションは、別の片方の売上増加要因になると解釈できる。「この製品を購入している方はこんな製品を購入しています」という重ね買いを促すプロモーション手法は、マスメディアによるプロモーションでも機能すると考えられるので、メディアを通して他の製品との類似関係を消費者に訴求する事は、有用なプロモーションになり得る可能性がある。

本研究では、類似関係を消費者に訴求するプロモーションが、売上増加要因となり得るという状況証拠を集め、その背景になり得る数理モデルを提示した。類似関係を訴求するプロモーションは、値下げなどの直接的なプロモーションに比べると即効性も無く、効果も小さいかもしれない。しかし、類似という相互効果による波及効果が見込める魅力がある。

競争空間を用いた新製品の評価

第5章では、類似係数を用いて競争空間を構成し、この布置における消費者と製品の距離が製品選択確率を説明する事を示した。この結果は、競争空間がレコメンドシステムの新たな分析フレームとなる可能性を示唆している。類似係数を算式として処理する場合に比べ、空間に埋め込んだデータに対しては、空間統計を利用したり、潜在顧客の分布を確認したりと、多様な分析手法を適用できる。レコメンドシステムを構築するシステムエンジニアに対する示唆として、類似係数を用いた競争空間の有用性を示すことができた。

また、競争空間を用いた新製品の新規採択者の分布を時系列で評価した結果、新規採択者の類似関係は次第に強くなる傾向があり、これが売上パターンに影響するという知見を導いた。この結果は、狭い範囲に分布する同質的な顧客のみをターゲットにすると、プロモーションとしては効率的であるが、製品ライフサイクルを短期化させる懸念が高まる事を示唆している。本研究で得た知見は、ライフサイクルを考慮した顧客戦略を考えるマーケティングに対する示唆になると期待する。

6.4 今後の課題

最後に、本研究の限界と今後の課題について考察する。

繰り返し購買への適用

本研究では、消費者の異質性を捉えたモデルを提示できなかった。製品カテゴリーの類似関係の評価値に異質性を与えず、また消費者の選好構造の空間的表現においても、消費者の異質性を考慮していない。消費者の異質性を捉えるには、データの自由度が足りないという制約もあるが、本研究の評価軸とする製品カテゴリーの類似関係とは異なる概念の消費者のグループ化において、異質性を確認するという対応が考えられる。これまでの試行的実験では、そのような結果は見出せていないが、製品カテゴリーの類似関係がどのように形成されるかを知るためには、類似関係の変化を異質性として捉える視点も重要と考える。

繰り返し購買への適用

本研究では、新製品の新規採択者に注目して研究を行った。手法の構成上は、評価対象が繰り返し購買のある日用品であってもかまわないが、2回目以降の繰り返し購入をカウントせず、最初の新規購入のみを評価するというデータ処理を行った。このような単純化を行わず、繰り返し購買の回数を評価しようとする、類似係数が購買の有無の2値のプロファイルデータを対象とする指標であるために、数理モデルとしての拡張を必要とする。

購買回数の差異を考慮するには、プロファイルデータが従う分布を2項分布から多項分布に変えれば良いという見通しが持てる。そのような相関係数の算出の拡張を行えば、他の分析手続きに大きな変更は必要ないかもしれない。繰り返し購買への拡張は興味深く、手法の適用範囲が広がるため、ぜひとも取り組みたい課題である。

時系列分析への応用

本研究では、一定期間の購買履歴における消費者の製品選択の有無を評価し、時系列の変化を分析する場合には、異なる期間の製品選択のプロファイルデータを作成することで、その変化を分析するという手順を用いた。プロファイルデータを取り扱う因子分析において時系列分析が一般的でないように、本研究手法を時系列分析に拡張する事は容易ではなからう。とはいえ、購買履歴において時間変化は重要な要因であり、A製品を購入してからB製品を購入するパターンは、その逆よりも多いといった特徴を取り込んで分析する事への、実務的ニーズは高いと思われる。

限定的ではあるが、競争空間を用いた新製品の新規採択者の類似関係の変化において、非可逆的な潜在顧客の同質化という時系列評価に基づく特徴を確認した。時系列分析への拡張においては、一般的な記述を目指すことよりも、このような知見を積み上げて、消費者行動に沿ったカスタマイズを行う事が実際的な対応と考える。そのために、様々な市場における実証分析を行う事を、今後の課題としたい。

参考文献

- Aaker, David. A. (1996) , 'Building Strong Brands,' The Free Press, 邦訳: 陶山計介・小林哲・梅本春夫・石垣智徳 (1997), 『ブランド優位の戦略 顧客を創造する BI の開発と実践』, ダイヤモンド社.
- Aaker, David A. (2004) , 'Brand Portfolio Strategy,' Free Press, 邦訳: 阿久津聡 (2005), 『ブランド・ポートフォリオ戦略』, ダイヤモンド社.
- Aaker, David A. (2001) , 'Developing Business Strategies,' Wiley, 邦訳: 今枝 昌宏 (2002), 『戦略立案ハンドブック』, 東洋経済新報社.
- Adomavicius, Gediminas and Alexander Tuzhilin (2005) , "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," IEEE Transactions on Knowledge and Data Engineering, 17(6), 734-749.
- Aksoy, Lerzan, Paul N. Bloom, Nicholas H. Lurie and Bruce Cooil (2006) , "Should Recommendation Agents Think Like People?," Journal of Service Research, 8(4), 297-315.
- Albatineh, Ahmed N., Magdalena Niewiadomska-Bugaj and Daniel Mihalko (2006) , "On Similarity Indices and Correction for Chance Agreement," Journal of Classification, 23(2), 301-313.
- Ansari, Asim, Skander Essegaier and Rajeev Kohli (2000(1)) , "Internet Recommendation Systems," Journal of Marketing Research, 37(3), 363-375.
- Ansari, Asim, Kamel Jedidi, and Sharan Jagpal (2000(2)) , "A Hierarchical Bayesian Methodology for Treating Heterogeneity in Structural Equation Models," Marketing Science, 19(4), 328-347.
- Ansari, Asim, Kamel Jedidi (2002) , "Heterogeneous factor analysis models: A Bayesian approach", Psychometrika, 67(1), 49-77.
- Ariely, Dan, John G. Lynch, Jr. and Manuel Aparicio IV (2004) , "Learning by Collaborative and Individual-Based Recommendation Agents," Journal of Consumer Psychology, 14(1&2), 81-94.
- Bass, Frank M. (1969) , "A New Product Growth Model for Consumer Durables", Management Science, 15 , 215-227.
- Batagelj, V., and Bren, M. (1995) , "Comparing resemblance measures," Journal of Classification, 12, 73-90.
- Baulieu, F.B. (1989) , "A classification of presence/absence based dissimilarity coefficients," Journal of Classification, 6, 233-246.

- Bishop, Christopher M. (2006) , 'Pattern Recognition and Machine Learning,' Springer, 邦訳: 元田浩, 栗田多喜夫, 樋口知之, 松本裕治, 村田昇 (監訳)(2007-2008), 『パターン認識と機械学習 上, 下 —ベイズ理論による統計的予測』, シュプリンガー・ジャパン.
- Bodapati, Anand V. (2008) , "Recommendation Systems with Purchase Data," *Journal of Marketing Research*, 45(1),77-93.
- Breese, John S., David Heckerman and Carl Kadie (1998) , "Proceedings of the fourteenth Annual Conference on Uncertainty in Artificial Intelligence," 43-52.
- Bruyn, Arnaud De, John C. Liechty, Eelko K. R. E. Huizingh and Gary L. Lilien(2008) , "Offering Online Recommendations with Minimum Customer Input Through Conjoint-Based Decision Aids," *MARKETING SCIENCE*,27(3),443-460.
- Cohen, J. (1960) , "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, 20, 37-46.
- Collins, Allan M. and Elizabeth F. Loftus(1975) , "A spreading activation theory of semantic memory," *Psychological Review*, 82, 407-428
- Duarte, J.M., Santos, J.B., and Melo, L.C. (1999) , "Comparison of similarity coefficients based on RAPD markers in the common bean," *Genetics and Molecular Biology*, 22, 427-432.
- Dubien, Janice L. and William D. Warde (1979) , "Mathematical Comparison of the Members of an Infinite Family of Agglomerative Clustering Algorithms," *Canadian Journal of Statistics*,7,29-38.
- Fishbein, M. (1963) , "An investigation of relationships between beliefs about an object and the attitude toward that object,". *Human Relations*, 16, 233-240.
- Fleiss, J.L.(1975) , "Measuring agreement between two judges on the presence or absence of a trait," *Biometrics*, 31, 651-659.
- Goldberg, David, David Nichols, Brian M. Oki and Douglas Terry (1992) , "Using collaborative filtering to weave an information tapestry," *Communications of the ACM*,35(12),61-70.
- Goodman, L.A. (1974) , "Exploratory Latent Structure Analysis Using both Identifiable and Unidentifiable Models," *Biometrika*, 61, 215-231.
- Goodman, L.A. (1996) , "A single general method for the analysis of cross-classified data: Reconciliation and synthesis of some methods of Pearson, Yule, and Fisher, and also some methods of correspondence analysis and association analysis," *Journal of the American Statistical Association*, 91, 408-428.
- Gordon, AD(1999) , 'Classification, 2nd Edition,' *Monographs on Statistics and Applied Probability* 82, Chapman & Hall/CRC.
- Gower, J. C. and Pierre Legendre, (1986) , 'Metric and Euclidean properties of dissimilarity coefficients,' *Journal of classification*, 3(1), 5-48.
- Grover Rajiv and V. Srinivasan(1987) , "A Simultaneous Approach to Market Segmentation and Market Structuring," *Journal of Marketing Research*, 24(2), 139-153.

- Grover Rajiv and V. Srinivasan(1989) "AAn Approach for Tracking Within-Segment Shifts in Market Shares," *Journal of Marketing Research*, 26, 230-236.
- Häubl, Gerald and Valerie Trifts (2000) , "Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids," *MARKETING SCIENCE*, 19(No.1),4-21.
- Häubl, Gerald and Kyle B. Murray (2003) , "Preference Construction and Persistence in Digital Marketplaces: The Role of Electronic Recommendation Agents," *Journal of Consumer Psychology*,13(1&2),75-91.
- Hubert, L., and Arabie, P. (1985) , "Comparing partitions," *Journal of classification*, 2, 193-218.
- Iacobucci, Dawn, Phipps Arabie and Anand Bodapati (2000) , "Recommendation agents on the internet," *Journal of Interactive Marketing*,14(3),2-11.
- Janson, S., and Vegelius, J. (1981) , "Measures of ecological association," *Oecologia*, 49, 371-376.
- Kamakura, Wagner A. and Gary J. Russell(1989) , "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research*, 26(4), 379-390.
- Kamakura, Wagner A., Nyung-Do Kim and Jonathan Lee (1996) , "Modeling Preference and Structural Heterogeneity", *Marketing Science*, 15(2) 152-172.
- Kaufman, L., and Rousseeuw, P.J. (1990) , "Finding Groups in Data," New York, USA: John Wiley & Sons, Inc.
- Keller, Kevin Lane (1998) , 'Strategic Brand Management', Prentice Hall, 監訳 恩蔵直人, 『戦略的ブランド・マネジメント』, 東急エージェンシー.
- Krider, Robert E. and Charles B. Weinberg(1998) , "Competitive Dynamics and the Introduction of New Products: The Motion Picture Timing Game," *Journal of Marketing Research*, 35(1), 1-5.
- Lance, G.N. and W.T. Williams (1967) , "A general theory of classificatory sorting strategies, 1.Hierarchical systems," *The Computer Journal*,9,373-380.
- Larkey, Levi B. and Arthur B. Markman(2005) "Processes of Similarity Judgment," *Cognitive Science*, 29(6), 1061-1076.
- Lehmann, Donald and Charles B. Weinberg(2000) , "Sales Through Sequential Distribution Channels: An Application to Movies and Videos." *Journal of Marketing*, 64 (July), 18-33.
- Linden, Greg, Brent Smith and Jeremy York (2003) , "Amazon.com recommendations: item-to-item collaborative filtering," *IEEE Internet Computing*,7 (No.1),76-80.
- Lurie, Nicholas H. (2004) , "Decision Making in Information-Rich Environments: The Role of Information Structure," *Journal of Consumer Research*,30(4),473-486.
- Manchanda, Puneet, Asim Ansari and Sunil Gupta (1999) , "The "Shopping Basket": A Model for Multicategory Purchase Incidence Decisions," *Marketing Science*, 18(2), 95-114.

- Mandel, J. (1971) , "A New Analysis of Variance Model for Non-additive Data," *Technometrics*, 13, 1, 1-18.
- Meyers-Levy, Joan and Alice M. Tybout (1989) , "Schema Congruity as a Basis for Product Evaluation," *Journal of Consumer Research*, 16 (June), 39-53.
- Milligan, G.W. (1996) , "Clustering validation: Results & implications for applied analyses," In *Clustering and Classification* (eds P. Arabie, L. J. Hubert & G. De Soete), World Scientific, Singapore, 341-375.
- Milligan, G.W. and Cooper, M.C. (1986) , "A Study of the Comparability of External Criteria for Hierarchical Cluster Analysis," *Multivariate Behavioral Research*, 21, 4, 441-458.
- Moe, Wendy W. and Peter S. Fader. (2001) , "Modeling Hedonic Portfolio Products: A Joint Segmentation Analysis of Music CD Sales," *Journal of Marketing Research*, 38(3), 376-385.
- Moon, Sangkil and Gary J. Russell (2005) , "Predicting product purchase from inferred customer similarity: An autologistic model approach," in *Proceedings of Tsukuba-Tohoku Joint Workshop on New Directions of Research in Marketing*, Tsukuba, Japan, December 2006.
- Punj, Girish N. and David W. Stewart (1983) "Cluster Analysis in Marketing Research: Review and Suggestions for Application," *Journal of Marketing Research*, 20, 134-48.
- Rand, W.M. (1971) , "Objective criteria for the evaluation of clustering methods," *Journal of the American Statistical Association*, 66, 336, 846-850.
- Ratneshwar, S and Allan D. Shocker(1991) "Substitution in use and the role of usage context in product category structures," *Journal of Marketing Research*, 28, 291-295.
- Resnick, Paul, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom and John Riedl (1994) , "GroupLens: an open architecture for collaborative filtering of netnews," *Proceedings of the 1994 International ACM Conference on Computer Supported Collaborative Work Conference*, 175-186.
- Rogers, E. M. (1983) , *Diffusion of Inovations* 3rd ed.', The Free Press, New York.
- Rosch, Eleanor (1975) , "Cognitive Reference Points", *Cognitive Psychology*, 7, 532-547.
- Rossi, Peter E., Greg M. Allenby, and Rob McCulloch (2006) , "Bayesian Statistics and Marketing", Wiley.
- Sawhney , Mohanbir S. and Jehoshua Eliashberg (1996) , "A Parsimonious Model for Forecasting Gross Box-Office Revenues of Motion Pictures, " *Marketing Science*, 15(2), 1996, 113-131.
- Shawe-Taylor, John and Nello Cristianini. (2004) , 'Kernel Methods for Pattern Analysis,' Cambridge University Press, 邦訳： 大北剛訳 (2010), 『カーネル法によるパターン解析』, 共立出版.
- Shocker, Allan D., Moshe Ben-Akiva, Bruno Boccara and Prakash Nedungadi(1991) , "Consideration Set Influences on Consumer Decision-Making and Choice: Issues, Models, and Suggestions," *Marketing Letters*, 2(3), 181-197.

- Sujan, Mita (1985) , "Consumer Knowledge: Effects on Evaluation Processes Mediating Consumer Judgments," *Journal of Consumer Research*, 12, (Jun), 31-46.
- Swaminathan, Vanitha (2003) , "The Impact of Recommendation Agents on Consumer Evaluation and Choice: The Moderating Role of Category Risk, Product Complexity, and Consumer Knowledge," *Journal of Consumer Psychology*, 13(1&2), 93-101.
- Tversky, Amos (1972) , "Elimination by aspects: A theory of choice", *Psychological Review*, 79(4), 281-299.
- Tversky, Amos (1977) , "Features of Similarity", *Psychological Review*, 84(4), 327-352.
- Urban, Glen L., Philip L. Johnson, and John R. Hauser(1984) , "Testing Competitive Market Structure," *Marketing Science*, 3(2), 83-112.
- Warrens, M.J. (2008) , "On the indeterminacy of resemblance measures for (presence/absence) data," *Journal of Classification*, 25, 125-136.
- Wedel, Michel and Wagner A. Kamakula(2000) , 'Market Segmentation: Conceptual and Methodological Foundations 2nd ed.' , Kluwer Academic Publisher.
- Yadohisa, Hiroshi, Takeuchi, Akinobu and Inada, Koichi(1999) , "Developing criteria for measuring space distortion in combinatorial cluster analysis and methods for controlling the distortion", *Journal of Classification*, 16(1), 45-62.
- 阿部 誠 (2003) , 「プロダクト空間とブランド空間を考慮したジョイント・スペース・マップ 北米ピックアップ・トラック市場への応用」 , オペレーションズ・リサーチ: 経営の科学 48(10), 719-728.
- 阿部 誠・近藤 文代 (2003) , 『マーケティング予測と発見科学 POS データの解析 - 』 , 朝倉書店.
- 荒木 長照 (2009) , 「劇場映画普及のミクロ的基礎とマクロ普及関数-moviegoer の観賞意思決定モデル-」 日本マーケティング・サイエンス学会 第 86 回研究大会.
- 石田 実, 西尾 チヅル, 佐藤 忠彦 (2008) , 「交互作用距離による音楽 CD の購買予測」 , 日本マーケティング・サイエンス学会 第 84 回研究大会プロジェクト研究報告 b-6.
- 石田 実・西尾 チヅル・佐藤 忠彦 (2009) , 「同一製品の繰返し購買が無い市場の新製品の属性に対して協調フィルタリングを用いる競争市場構造分析手法の提案」 , 筑波大学 企業科学専攻 リサーチレポート, 09-01.
- 石田 実, 西尾 チヅル, 椿 広計 (2011) , 「2 値変量に基づく教師無し分類における類似係数の選択」 *日本行動計量学会和文誌*, 38(1), 65-81.
- 一般社団法人 日本レコード協会 (2008,2009) , 「音楽メディアユーザー実態調査」 , <http://www.riaj.or.jp/report/mediauser/index.html>.
- 井上 哲浩・中西 正雄 (1990) , 「異質性を組み入れた競争市場構造分析」 , マーケティング・サイエンス, 35, 9-17.
- 井上 哲浩 (1992) , 「異質性を伴った競合グループ識別モデルの一システム:MIGHT」 , マーケティング・サイエンス, 1(1&2)12-37 .

- 井上 哲浩 (2001), 「競争市場構造分析」, 朝倉書店, 編者 岡太 彬訓, 木島 正明, 守口 剛, 『マーケティングの数理モデル』.
- 大隈 昇・L. ルバール・A. モリナウ・K.M. ワーウィック・馬場康維 (1994), 『記述的多変量解析法』, 日科技連出版社.
- 片平 秀貴 (1987), 『マーケティング・サイエンス』, 東京大学出版会.
- 勝又 壮太郎, 阿部 誠 (2007), 「顧客嗜好の時間的変化を組み込んだ音楽CD選考モデルの構築とCRMへの応用」, マーケティング・サイエンス, 16(1-2), 25-47.
- 神竄 敏弘 (2007), 「推薦システムのアルゴリズム (1)」, 人工知能学会誌, 22(6), 826-837.
- 神竄 敏弘 (2008(1)), 「推薦システムのアルゴリズム (2)」, 人工知能学会誌, 23(1), 89-103.
- 神竄 敏弘 (2008(2)), 「推薦システムのアルゴリズム (3)」, 人工知能学会誌, 23(2), 248-263.
- 佐藤忠彦 (2010), 「マーケティングにおける結果データの動的活用のためのページアンモデリング」, オペレーションズ・リサーチ, 1月号, 25-30, 2010.
- 清水 聡 (1999), 『新しい消費者行動』, 千倉書房.
- 高橋 広行 (2011), 『カテゴリーの役割と構造 ブランドとライフスタイルをつなぐもの』, 関西学院大学出版会.
- 照井 伸彦 (2008), 『ベイズモデリングによるマーケティング分析』, 東京電機大学出版局.
- 中西 正雄 (1984), 「消費者情報処理理論」, 誠文堂新光社, 編者 中西正雄, 『消費者行動分析のニュー・フロンティア: 多属性分析を中心に』.
- 西尾 チヅル・竹内 淑恵 (2011), 「日本型 LOHAS 消費者の特質とその動向」, 日本マーケティング・サイエンス学会 第 89 回研究大会プロジェクト研究報告 b-6.
- 新倉 貴士 (2005), 『消費者の認知世界: ブランドマーケティング・パースペクティブ』, 千倉書房.
- 濱岡 豊 (1993), 「消費者間相互依存/相互作用」, マーケティング・サイエンス, 2, 60-85.
- 芳賀 麻誉美 (2008), 「レコメンデーション技術の理論と実際」, 日本行動計量学会第 36 回大会チュートリアルセミナー.
- 樋口 知之 (2011), 『予測にいかす統計モデリングの基本 ベイズ統計入門から応用まで』, 編著 繁桝 算男・森 敏昭・柳井 晴夫, サイエンス社.
- 広津 千尋 (1992), 『実験データの解析 - 分散分析を超えて』, 共立出版.
- 前川 真一 (1999), 『Q&A で知る統計解析 - DOs and DON'Ts』, 講談社.
- 柳井 晴夫, 前川 真一, 繁桝 算男, 市川 雅教 (1990), 『因子分析 その理論と方法』, 朝倉書店, 7 章.
- 山本 昭二 (1999), 『サービス・クォリティ - サービス品質の評価過程 - 』, 千倉書房.

渡辺 洋 (2008) , 『人文社会科学の統計学』, 東京大学出版会, 東京大学教養学部統計学教室, 第 11 章.

謝辞

西尾チヅル先生には、修士課程に入学した2000年より10年あまりの間、主指導教官としてマーケティングの初歩から指導頂きました。それまで学問としても実務としてもマーケティングに接した事のない私が、なんとか論文をまとめられるまでに至ったのは、西尾先生の長きにわたる御指導のおかげです。また、しばしば研究発表の機会を頂いたことは、研究を継続する大きなモチベーションとなりました。このようなご配慮がなかったら、既に論文執筆を投げ出していたでしょう。感謝申し上げます。

椿広計先生からは、交互作用統計量について考えるきっかけを頂きました。当初、交互作用統計量は単なる思いつきとして、椿先生の授業で自由発表した分析指標にすぎませんでした。後日に椿先生からコメントを頂いた事が励みとなり、また関連する著作を紹介して頂いたことから見識が深まり、交互作用統計量の優位性に確信を持ち、研究に取り組むことができました。

佐藤忠彦先生からは研究を進める途上で様々なコメントを頂戴し、見落としていた課題と対応のアイデアを示して頂きました。また、ともすれば数理的な分析のみに偏りがちな筆者に対し、実務への応用の重要性を忍耐強く指導していただきました。これから筆者がマーケティング上の応用に留意して研究を行う事ができたとしたら、佐藤先生のおかげによるものです。

学外の諸先生からも、ご指導や有益なコメントを多数頂戴しました。特に、浅野熙彦先生、阿部誠先生、井上哲浩先生、片平秀貴先生、中西正雄先生からは、学会発表の際に貴重なコメントや研究のアイデアを頂戴しました。

匿名の2人の査読者からは、研究の構成上の課題に対しての指導的、教育的な指摘を多数頂戴しました。査読者の指摘により、Tverskyの対比モデルの研究の存在に気付くことができ、より広い視野を持って研究の位置づけを見直すことができました。

西尾研究室の皆様からのコメントや会話は、大きな助けとなりました。例えば、研究手法が定まらない頃、普及理論や潜在クラス分析の周辺で迷子になり、西尾先生との個人ゼミで

進捗を示すこともできなかった頃、西尾先生と駒走さんが私の研究についてコメントするのを伺い、精緻なモデルを扱う前に類似係数について理解することが大切と気づき、研究の方針を定めることができました。

当然ながら、本研究は引用した多くの文献に依存しています。面識のない先人達に心から感謝を申し上げます。彼らと同様に、同じテーマに取り組む研究者と知見を分かち合えられるように精進することで、感謝の意を示したいと思います。最後にデータを提供頂いた匿名の企業と担当者様に、深謝申し上げます。

付 録 A

A.1 第 3 章の分析結果一覧

図 A-1-1 デンドログラム（空間の歪みの分類）

製品/消費者 (2) × 階層的クラスター法 (4) × 類似度 (16)

図 A-1-2 クラスターサイズ分布の交差確認（均等性）

製品/消費者 (2) × 階層的クラスター法 (4) × 類似度 (16)

2-fold 交差確認 × 50 回 = 100 通りの分類に対する boxplot

表 A-1-1 クラスターの大きさ上位半分の全体に対する比率（均等性）

本論の表 5 の詳細データ、クラスター法 (5) 別 10,15,20,25,30 分類毎の集計

表 A-1-2 再現性評価

本論の表 6 の詳細データ、階層的クラスター法 (4) 別 10,15,20,25,30 分類毎の集計

表 A-1-3 外的基準との独立性検証

本論の表 9 の詳細データ、クラスター法 (5) 別 10,15,20,25,30 分類毎の集計

表 A-1-4 消費者分類 × 製品分類の度数表における独立性検証（内的基準）

本論の表 10 の詳細データ、クラスター法 (5) 別 10,15,20,25,30 分類毎の集計

表 A-1-5 アーティスト名一覧（解釈可能性）

交互作用統計量 × ward 法で 15 分類別、販売度数順

図 A-1-1. デンドログラム -1

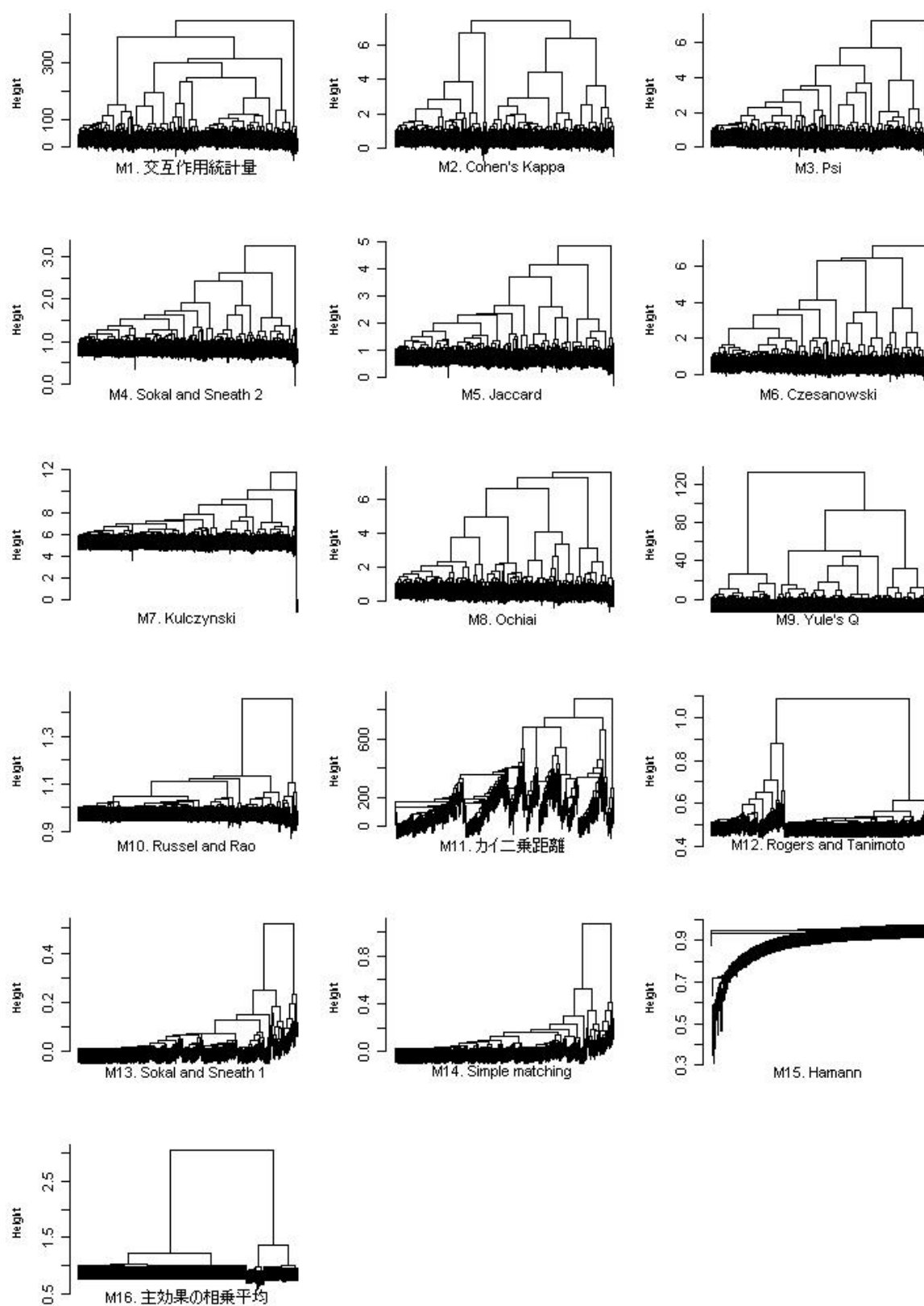


図 A-1-1-A-1. 製品分類 × 1.ward 法

図 A-1-1. デンドログラム -2

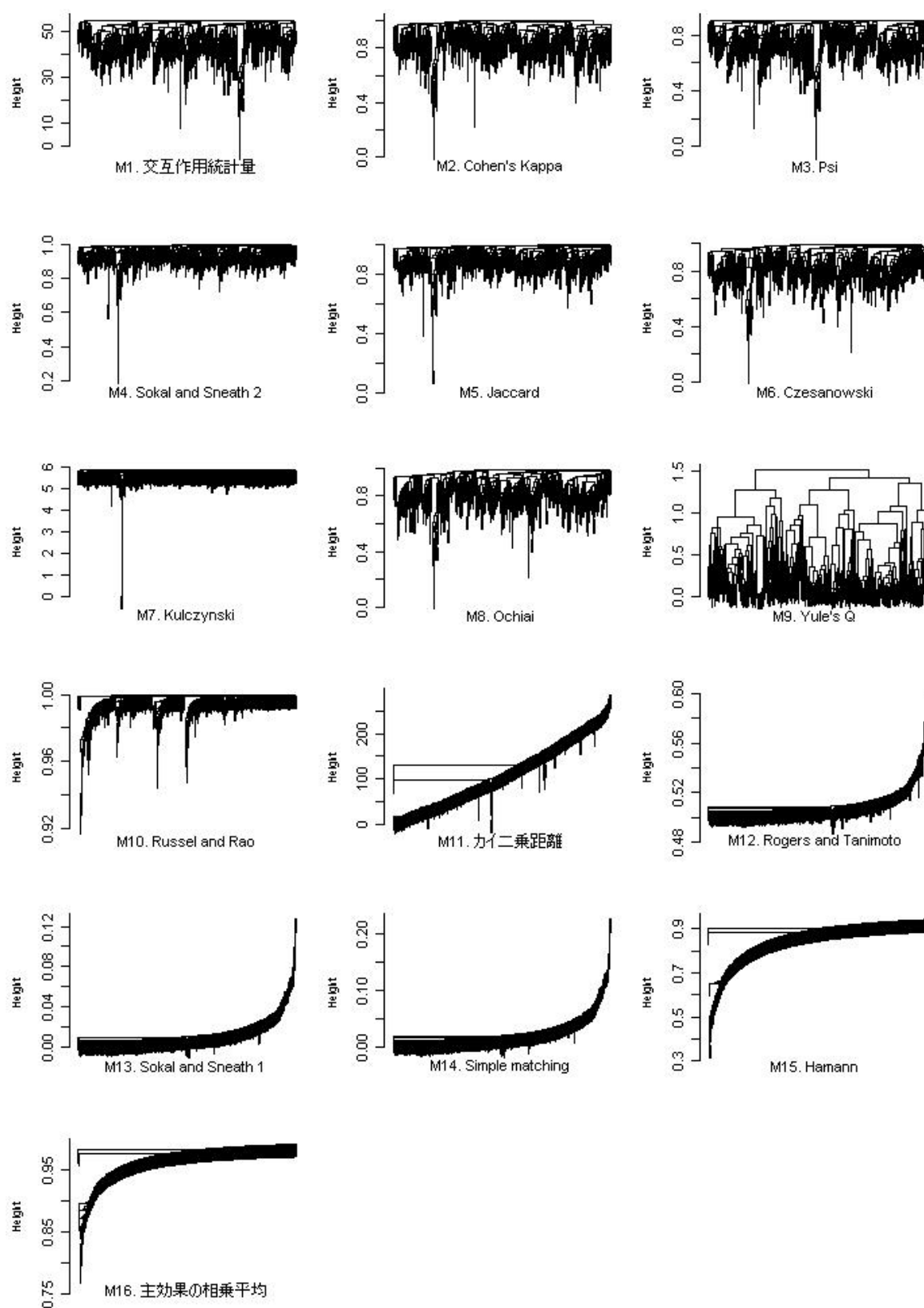


図 A-1-1-A-2. 製品分類 × 2. 群間平均法

図 A-1-1. デンドログラム -3

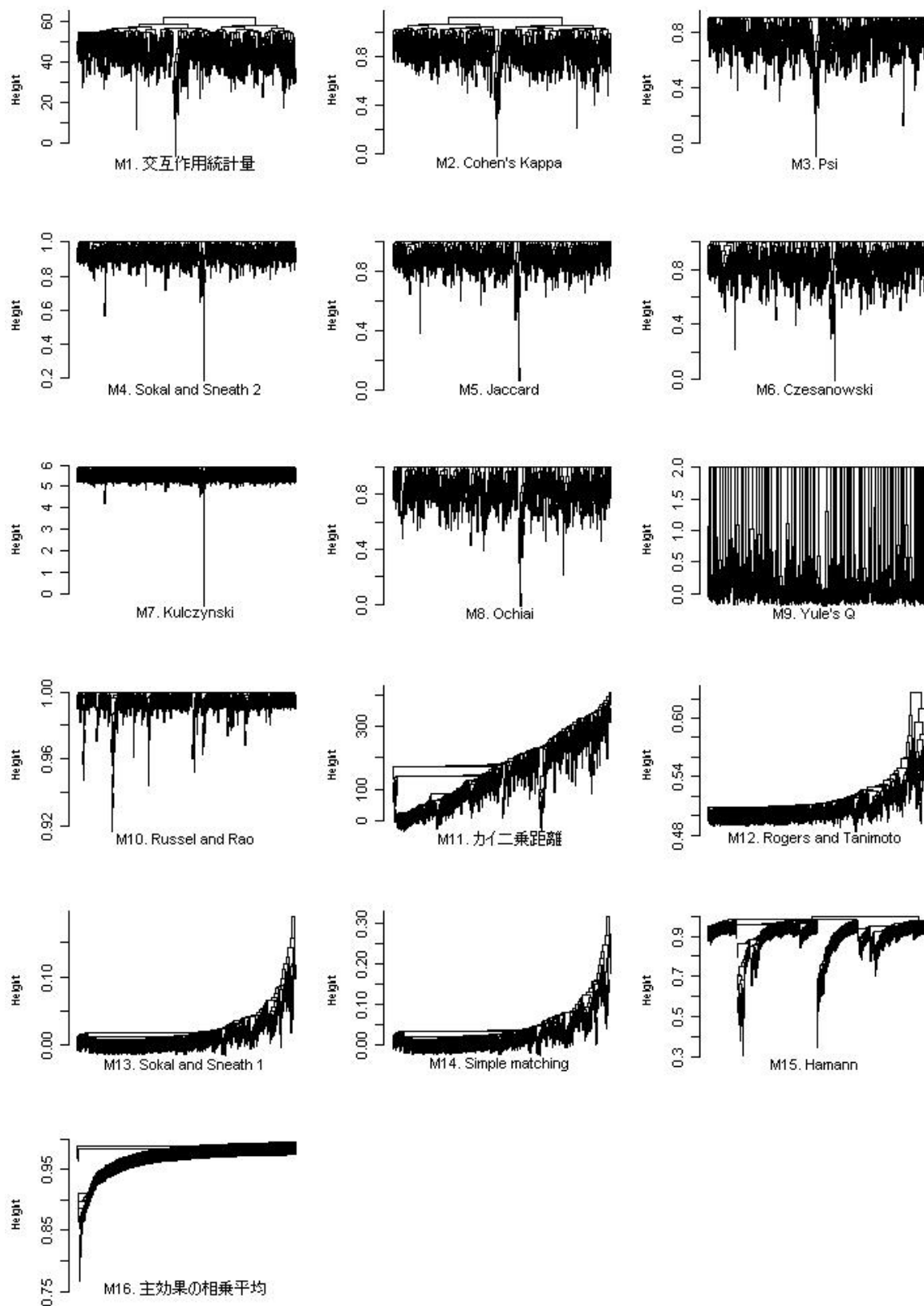


図 A-1-1-A-3. 製品分類 × 3. 最長距離法

図 A-1-1. デンドログラム -4

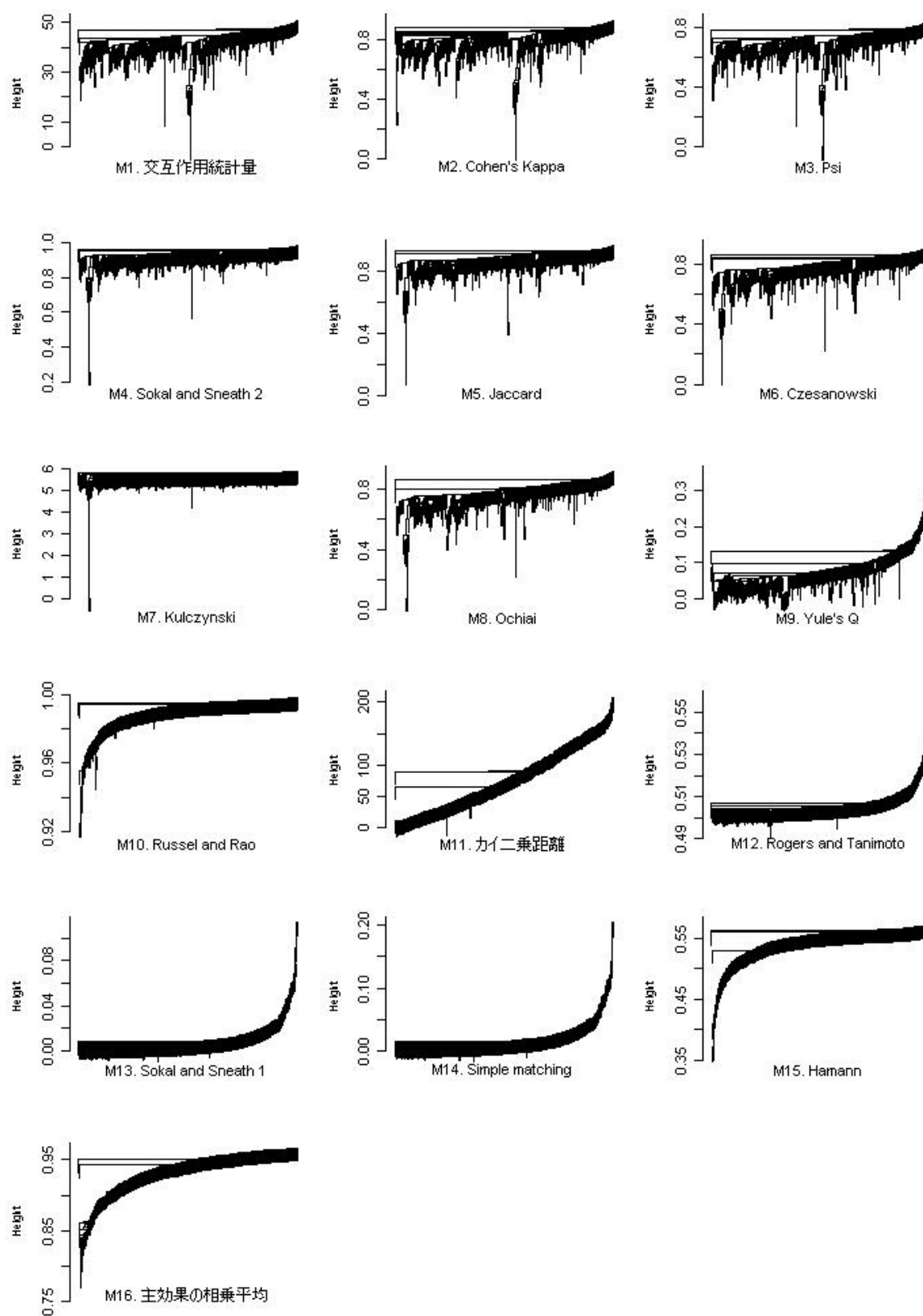


図 A-1-1-A-4. 製品分類 × 4. 最短距離法

図 A-1-1. デンドログラム -5

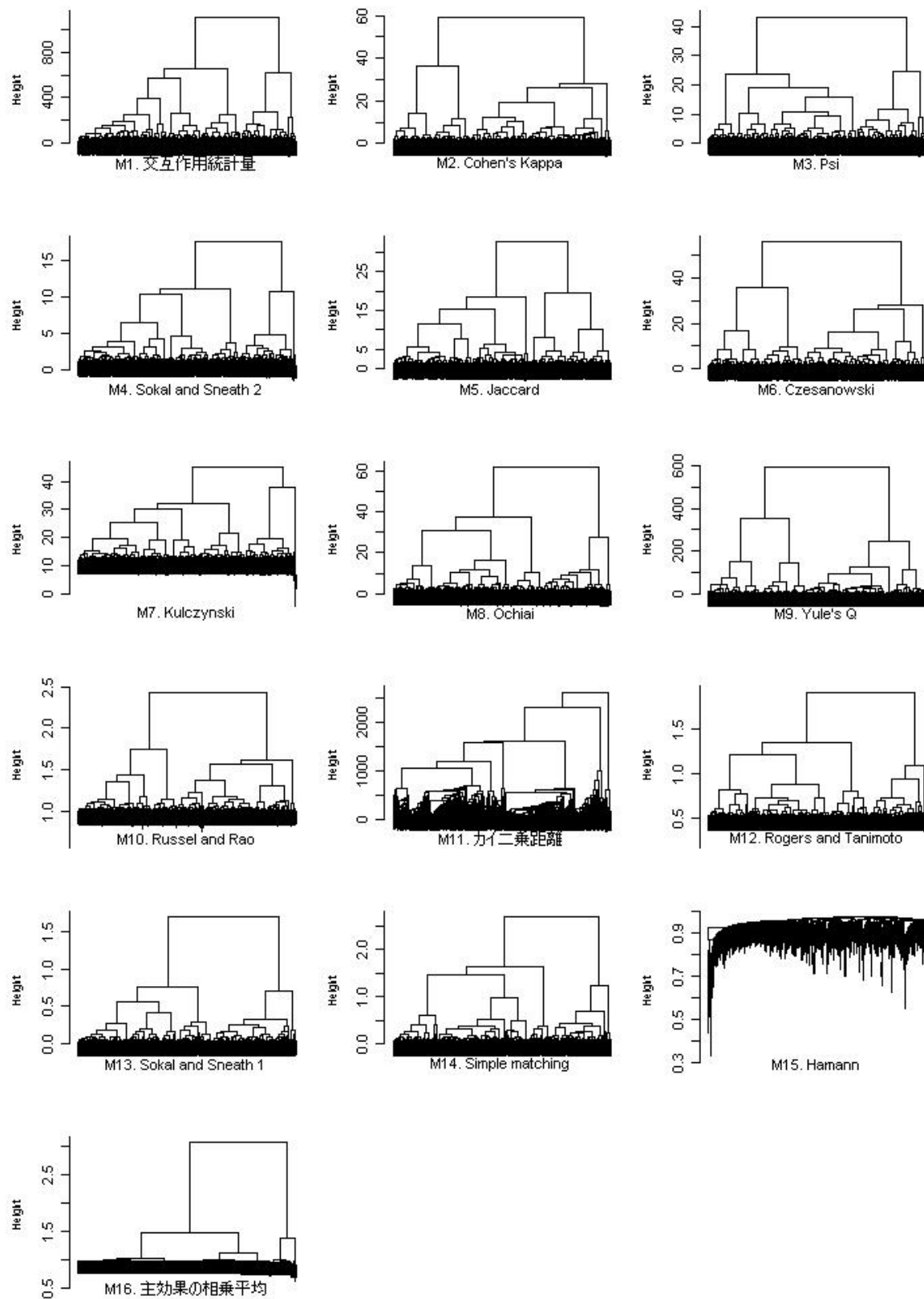


図 A-1-1-B-1. 消費者分類 × 1.ward 法

図 A-1-1. デンドログラム -6

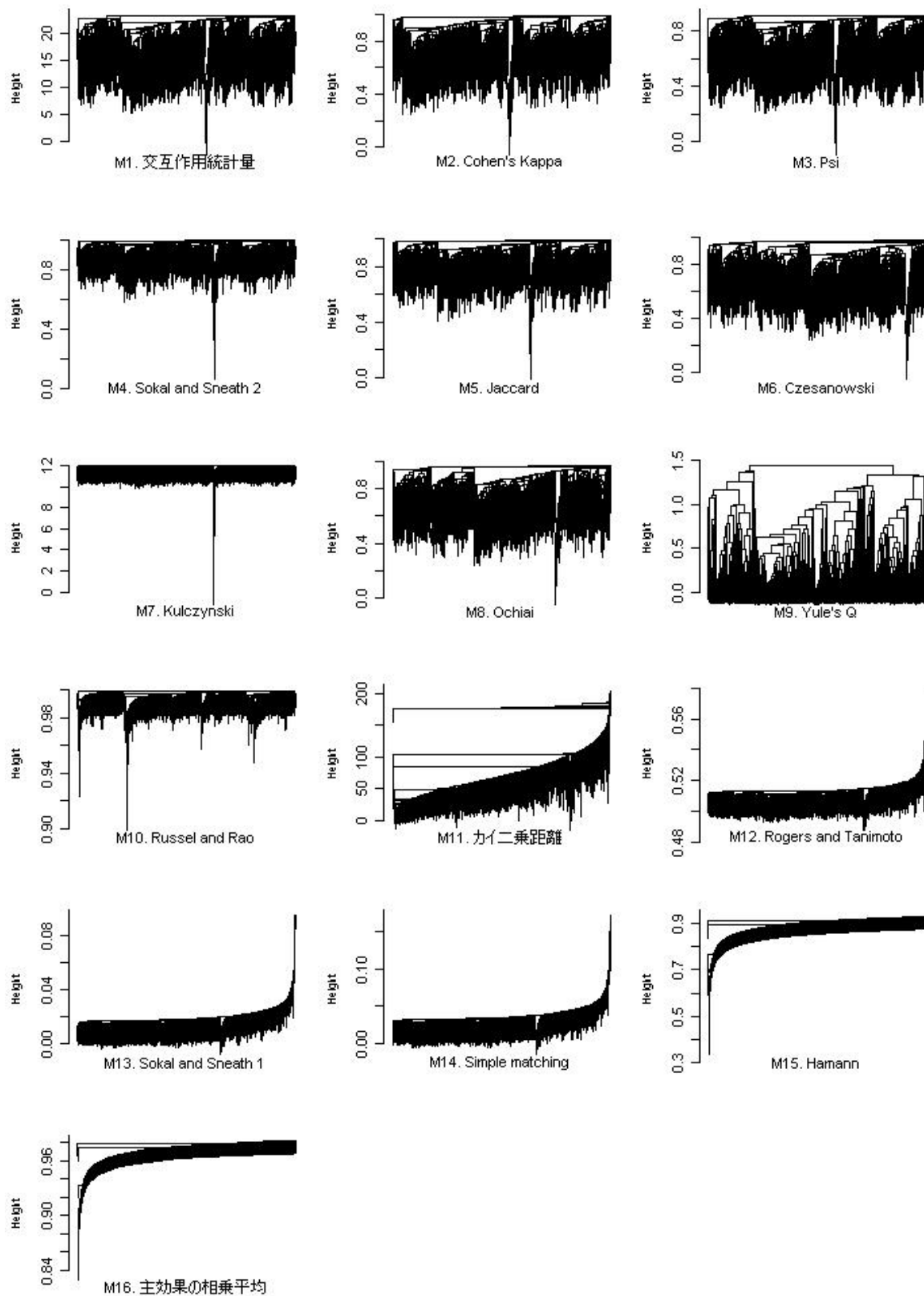


図 A-1-1-B-2. 消費者分類 × 2. 群間平均法

図 A-1-1. デンドログラム -7

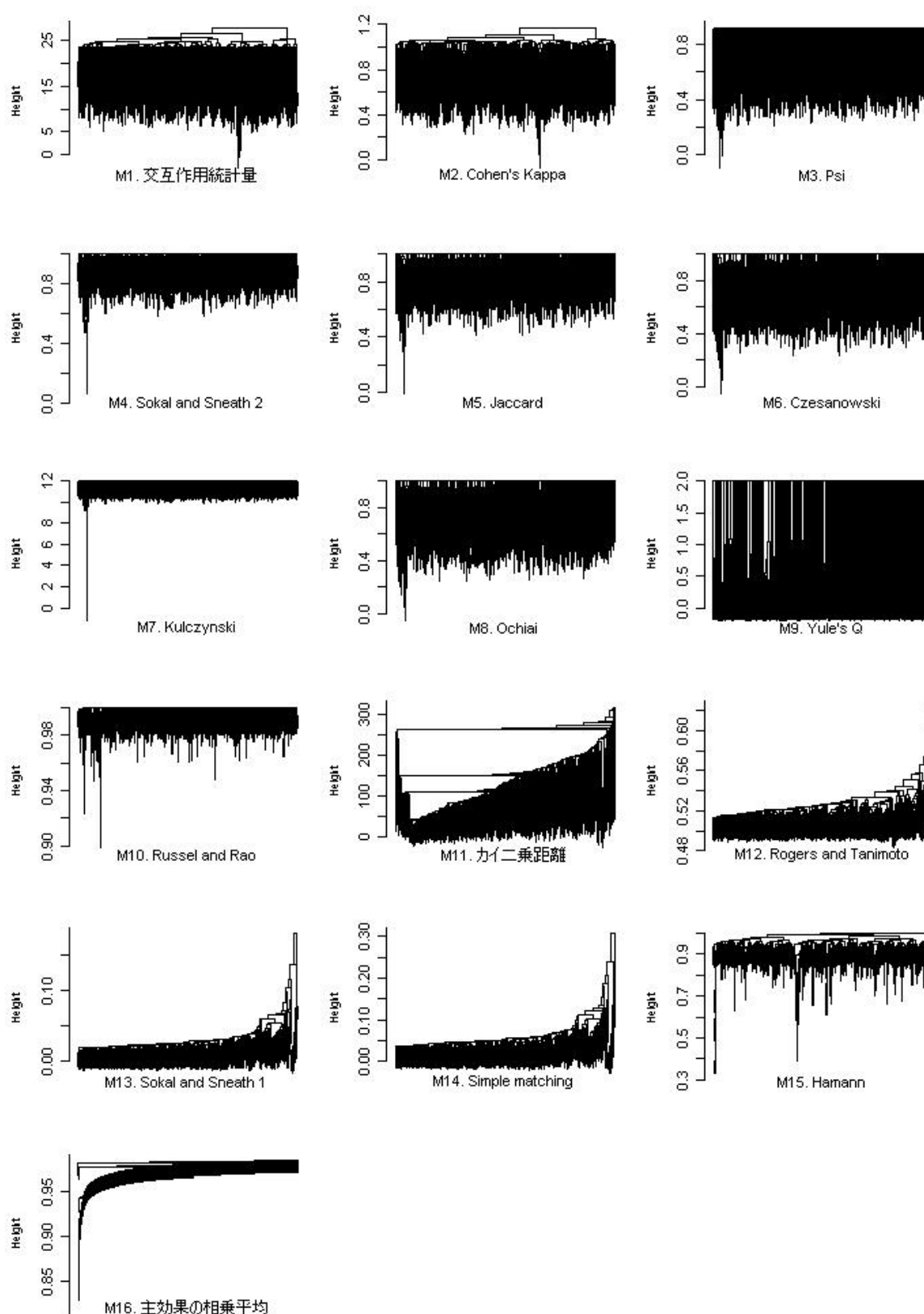


図 A-1-1-B-3. 消費者分類 × 3. 最長距離法

図 A-1-1. デンドログラム -8

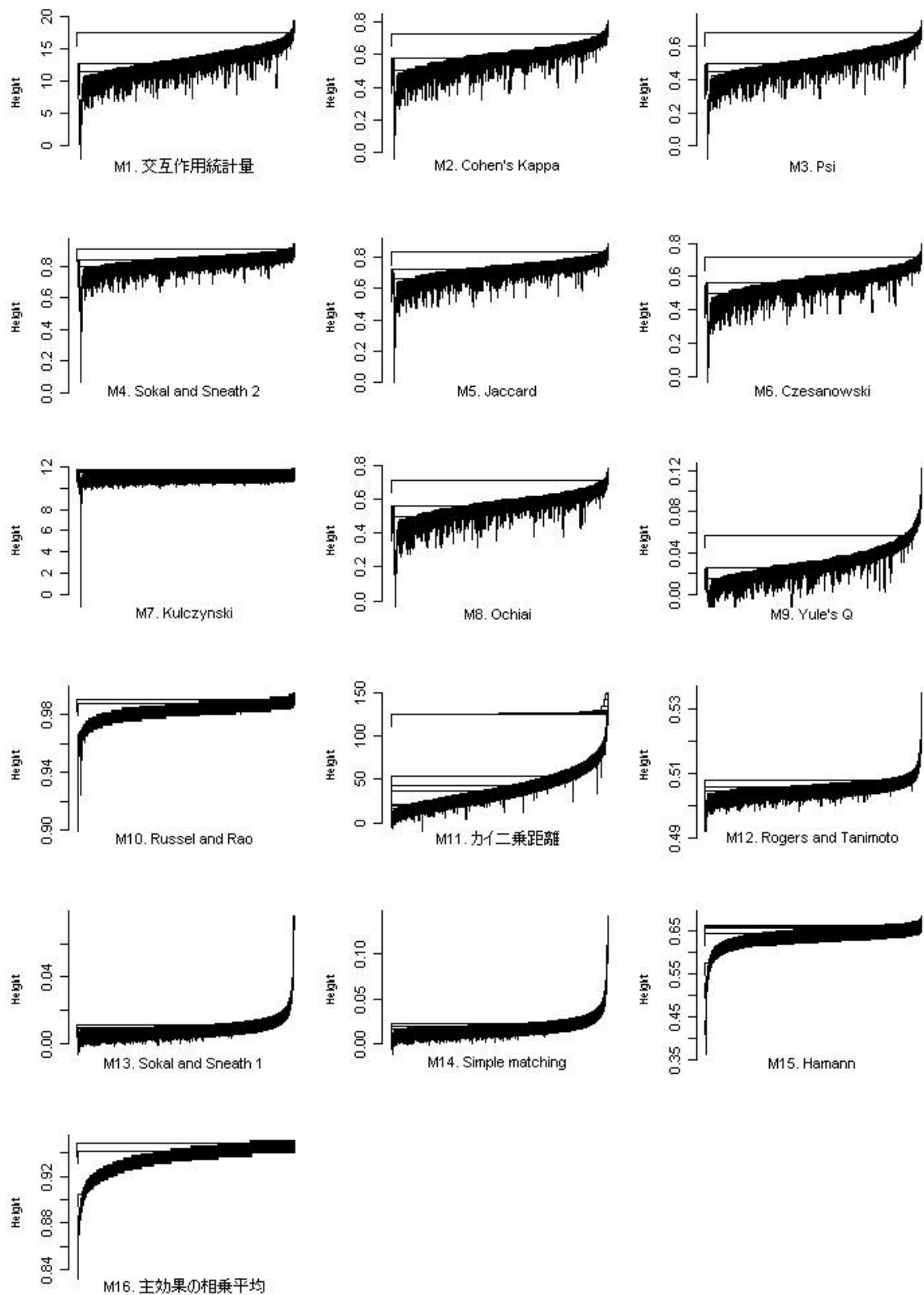


図 A-1-1-B-4. 消費者分類 × 4. 最短距離法

図 A-1-2. クラスターサイズ分布の交差確認 -1

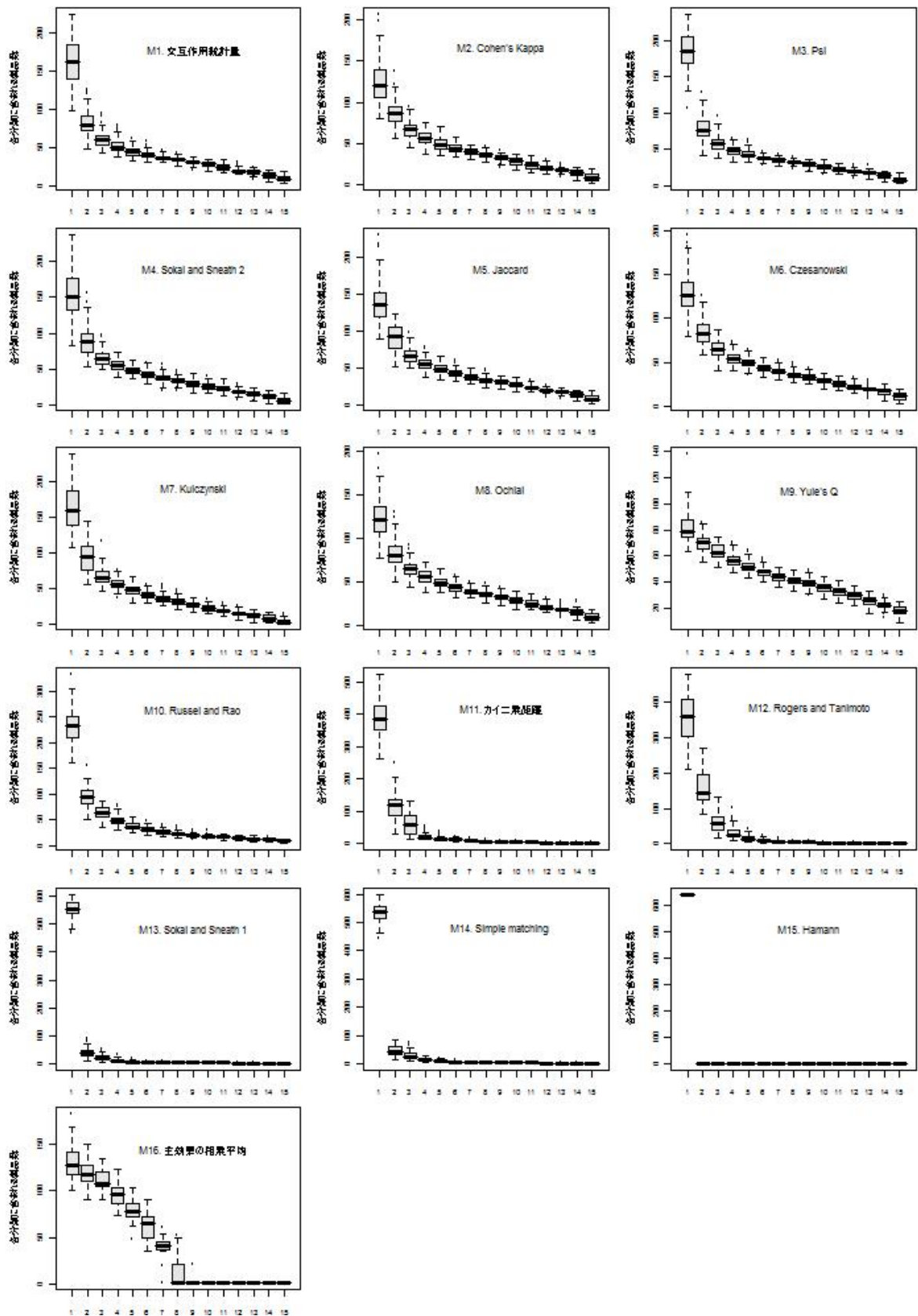


図 A-1-2-A-1. A. 製品分類 × 1.ward 法

図 A-1-2. クラスターサイズ分布の交差確認 -2

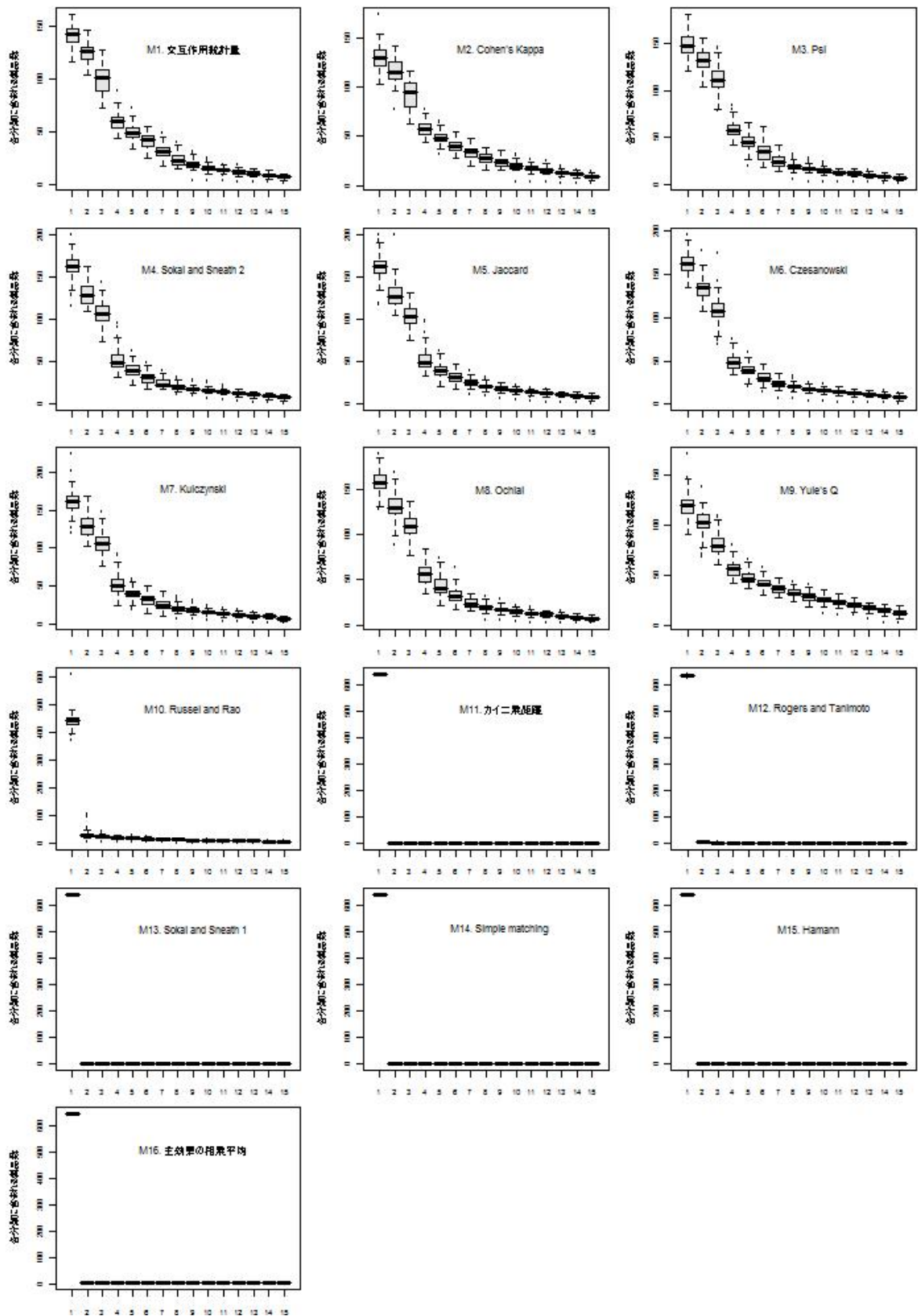


図 A-1-2-A-2. A. 製品分類 × 2. 群間平均法

図 A-1-2. クラスターサイズ分布の交差確認 -3

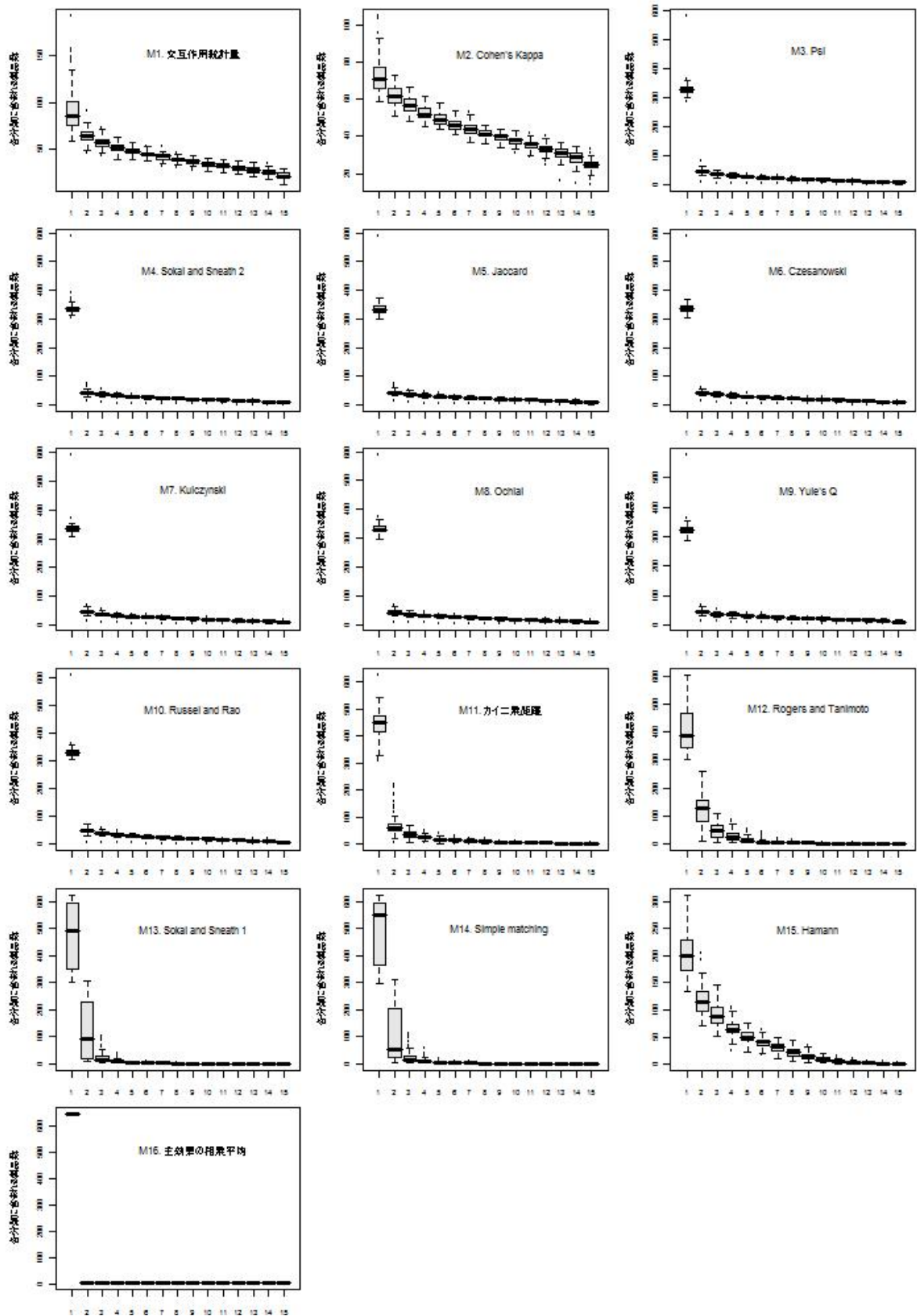


図 A-1-2-A-3. A. 製品分類 × 3. 最長距離法

図 A-1-2. クラスターサイズ分布の交差確認 -4

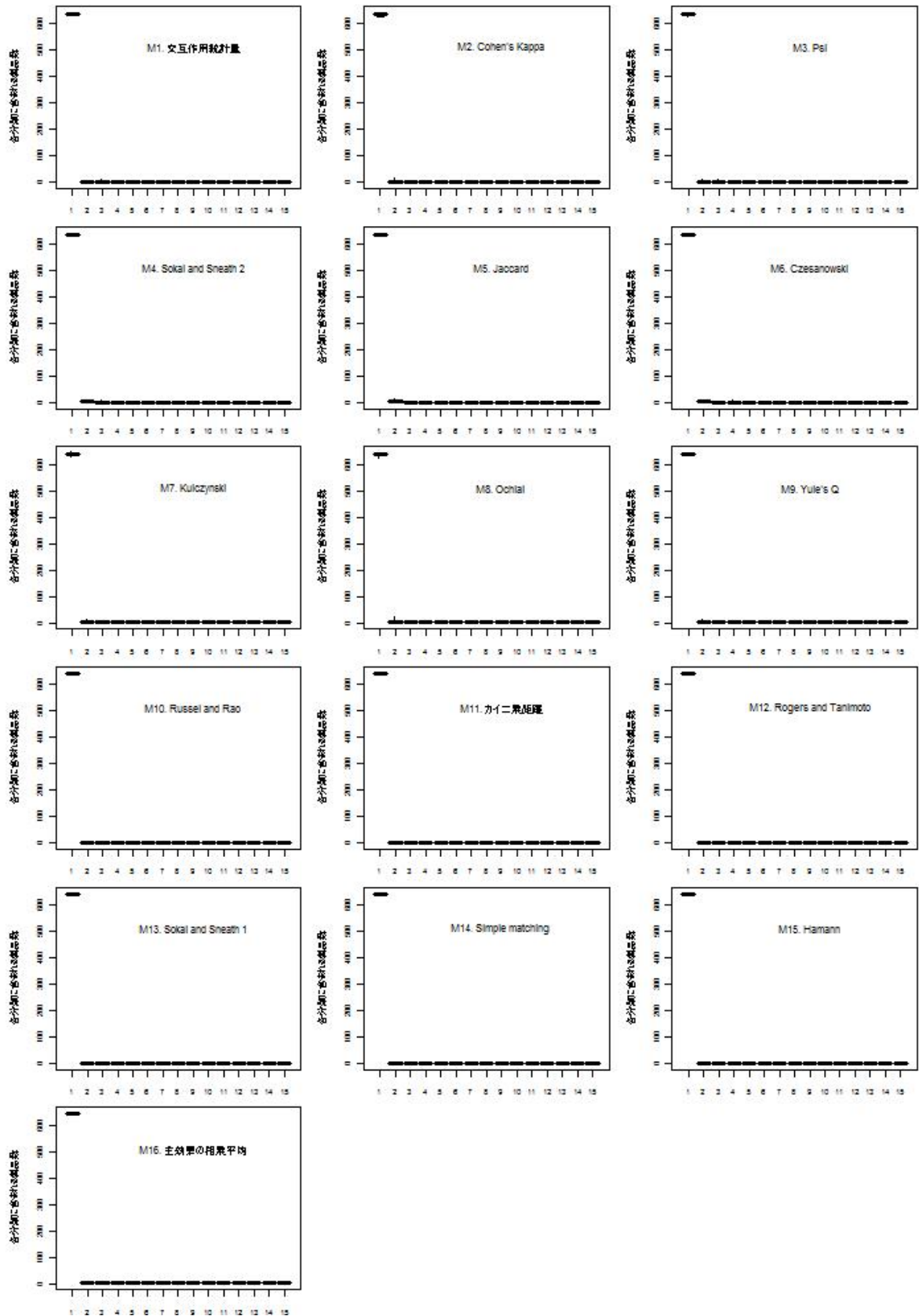


図 A-1-2-A-4. A. 製品分類 × 4. 最短距離法

図 A-1-2. クラスターサイズ分布の交差確認 -5

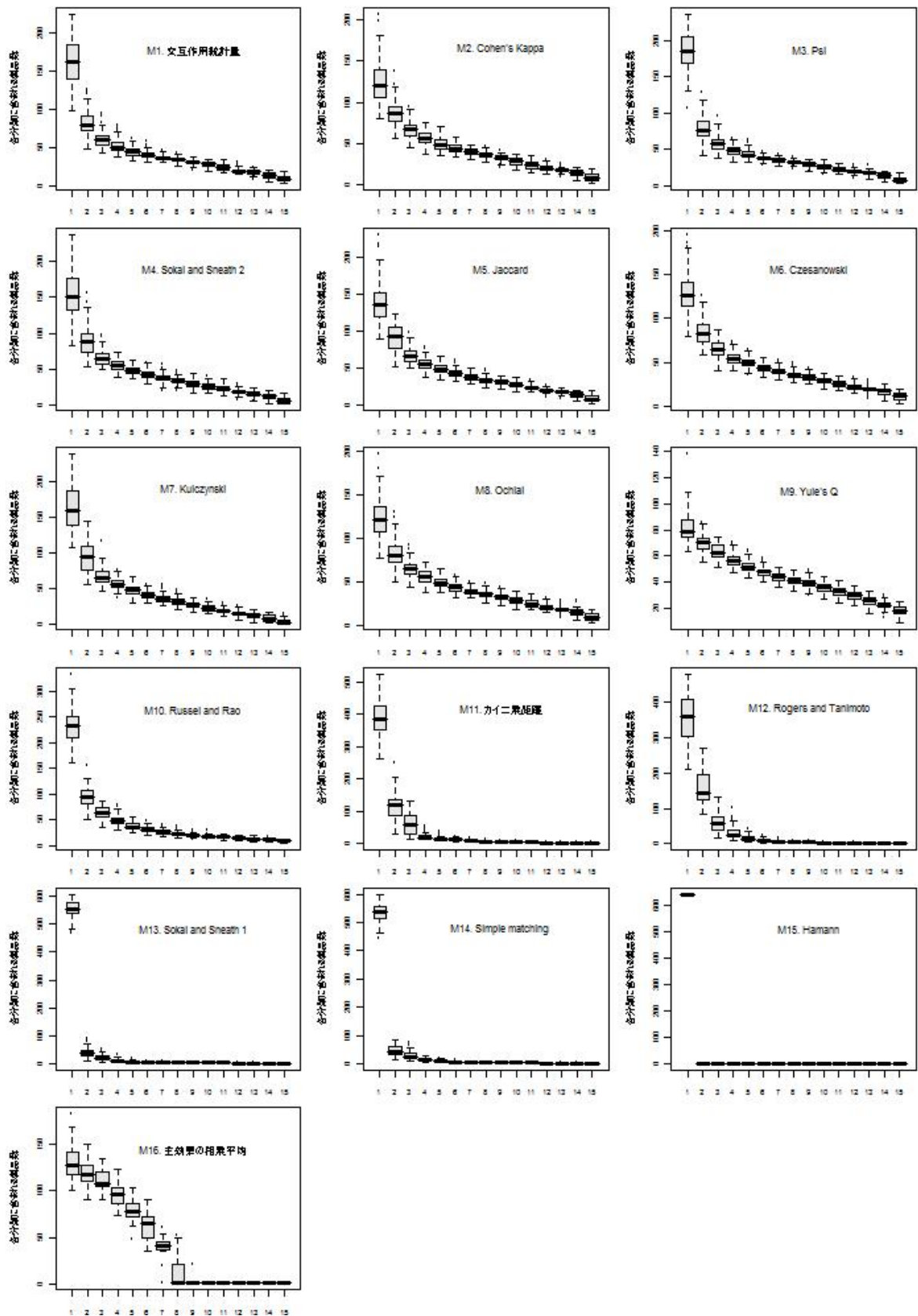


図 A-1-2-B-1. B. 消費者分類 × 1.ward 法 15 分類

図 A-1-2. クラスターサイズ分布の交差確認 -6

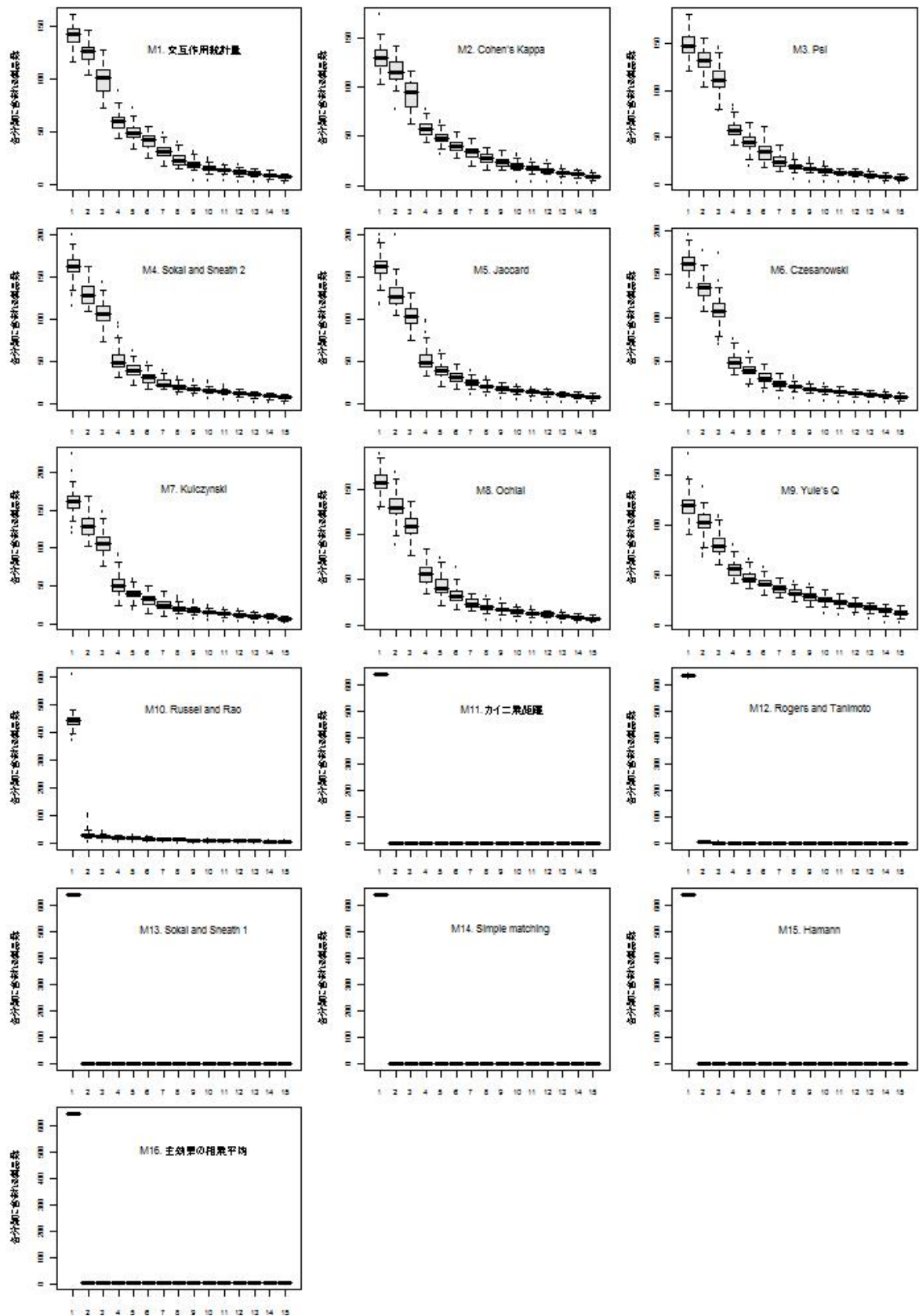


図 A-1-2-B-2. B. 消費者分類 × 2. 群間平均法 15 分類

図 A-1-2. クラスターサイズ分布の交差確認 -7

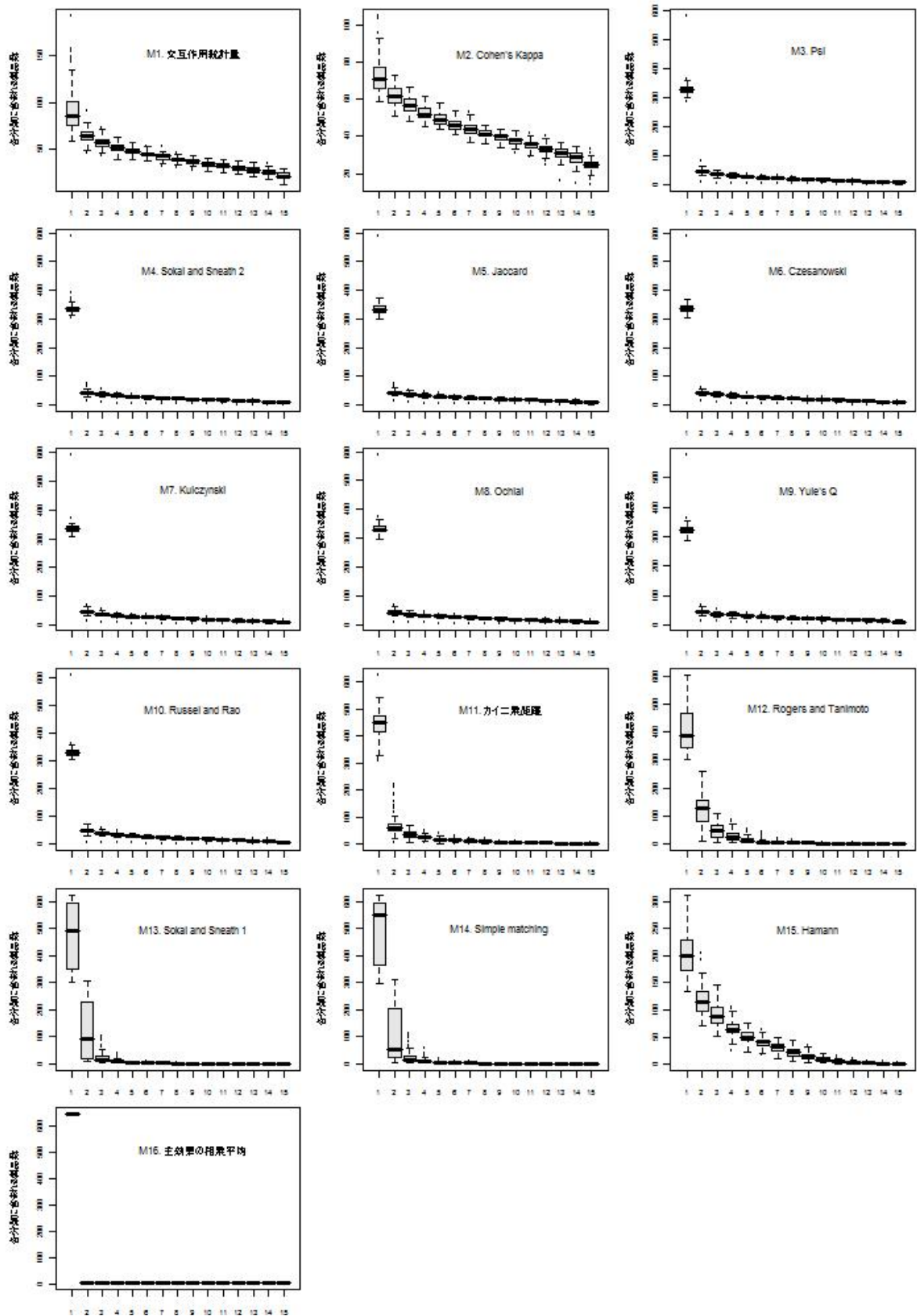


図 A-1-2-B-3. B. 消費者分類 × 3. 最長距離法 15 分類

図 A-1-2. クラスターサイズ分布の交差確認 -8

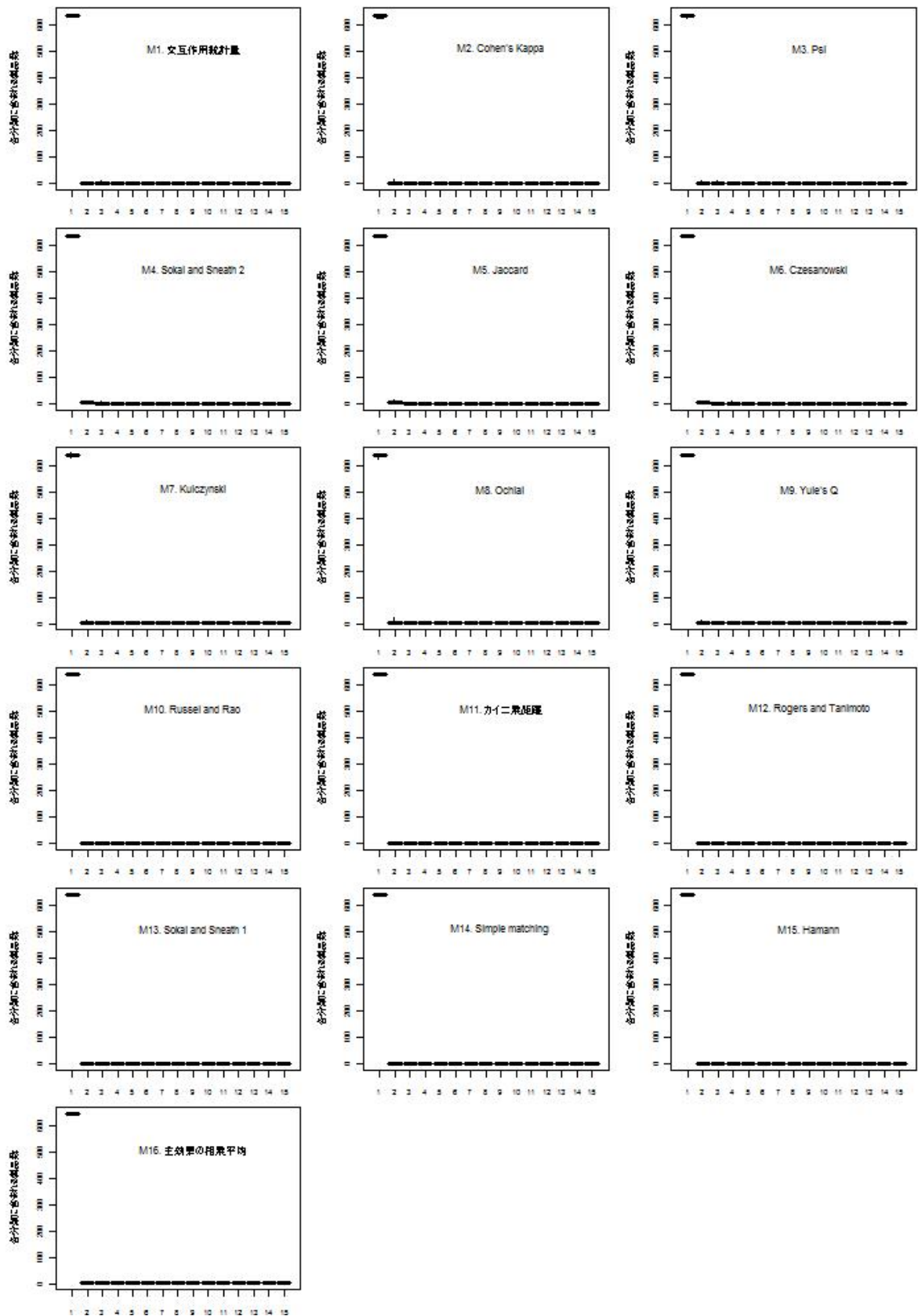


図 A-1-2-B-4. B. 消費者分類 × 4. 最短距離法 15 分類

表 A-1-1-1. クラスターの大きさ上位半分の全体に対する比率 - PAM 法

非類似度	10 分類	15 分類	20 分類	25 分類	30 分類
M1 交互作用統計量	67	66	68	65	64
M2 Cohen's Kappa	63	60	59	60	61
M3 Psi	67	67	68	65	64
M4 Sokal and Sneath 2	65	62	65	64	64
M5 Jaccard	65	62	63	63	62
M6 Czesanowski	67	63	63	62	62
M7 Kulczynski	72	67	66	65	65
M8 Ochiai	72	69	69	70	68
M9 Yule's Q	62	61	59	62	63
M10 Russel and Rao	73	77	86	92	95
M11 カイ二乗距離	99	99	98	98	98
M12 Rogers and Tanimoto	99	99	98	98	98
M13 Sokal and Sneath 1	99	99	98	98	98
M14 Simple matching	99	99	98	98	98
M15 Hamann	99	99	98	98	98
M16 主効果の相乗平均	99	99	98	98	98

表 A-1-1-2. クラスターの大きさ上位半分の全体に対する比率 - ward 法

非類似度	10 分類	15 分類	20 分類	25 分類	30 分類
M1 交互作用統計量	77	75	75	74	73
M2 Cohen's Kappa	76	75	72	72	71
M3 Psi	79	77	76	74	73
M4 Sokal and Sneath 2	79	78	77	77	75
M5 Jaccard	77	76	75	74	73
M6 Czesanowski	76	74	73	72	71
M7 Kulczynski	81	80	79	78	76
M8 Ochiai	74	73	72	71	70
M9 Yule's Q	64	66	67	68	69
M10 Russel and Rao	86	83	80	78	77
M11 カイ二乗距離	96	96	95	95	95
M12 Rogers and Tanimoto	97	97	97	97	97
M13 Sokal and Sneath 1	98	98	98	98	97
M14 Simple matching	98	98	98	98	97
M15 Hamann	99	99	98	98	98
M16 主効果の相乗平均	84	98	98	98	98

表 A-1-1-3. クラスターの大きさ上位半分の全体に対する比率 - 群間平均法

非類似度	10 分類	15 分類	20 分類	25 分類	30 分類
M1 交互作用統計量	84	85	85	84	84
M2 Cohen's Kappa	80	81	81	81	80
M3 Psi	88	86	85	84	84
M4 Sokal and Sneath 2	86	85	84	83	83
M5 Jaccard	86	85	84	83	83
M6 Czesanowski	86	85	84	83	83
M7 Kulczynski	86	85	84	83	83
M8 Ochiai	87	86	85	84	83
M9 Yule's Q	76	76	76	75	74
M10 Russel and Rao	91	89	87	87	86
M11 カイ二乗距離	99	99	98	98	98
M12 Rogers and Tanimoto	99	99	98	98	98
M13 Sokal and Sneath 1	99	99	98	98	98
M14 Simple matching	99	99	98	98	98
M15 Hamann	99	99	98	98	98
M16 主効果の相乗平均	99	99	98	98	98

表 A-1-1-4. クラスターの大きさ上位半分の全体に対する比率 - 最長距離法

非類似度	10 分類	15 分類	20 分類	25 分類	30 分類
M1 交互作用統計量	62	64	67	68	69
M2 Cohen's Kappa	60	62	63	64	65
M3 Psi	85	82	80	78	76
M4 Sokal and Sneath 2	85	83	80	78	76
M5 Jaccard	85	82	80	78	76
M6 Czesanowski	85	82	80	77	76
M7 Kulczynski	85	82	80	78	76
M8 Ochiai	85	82	79	77	75
M9 Yule's Q	83	81	78	76	74
M10 Russel and Rao	85	84	82	81	79
M11 カイ二乗距離	95	95	94	94	94
M12 Rogers and Tanimoto	97	97	98	98	97
M13 Sokal and Sneath 1	99	99	98	98	98
M14 Simple matching	99	99	98	98	98
M15 Hamann	81	93	97	98	97
M16 主効果の相乗平均	99	99	98	98	98

表 A-1-1-5. クラスターの大きさ上位半分の全体に対する比率 - 最短距離法

非類似度	10 分類	15 分類	20 分類	25 分類	30 分類
M1 交互作用統計量	99	99	98	98	98
M2 Cohen's Kappa	99	99	98	98	98
M3 Psi	99	99	98	98	98
M4 Sokal and Sneath 2	99	99	98	98	98
M5 Jaccard	99	99	98	98	98
M6 Czesanowski	99	99	98	98	98
M7 Kulczynski	99	99	98	98	98
M8 Ochiai	99	99	98	98	98
M9 Yule's Q	99	99	98	98	98
M10 Russel and Rao	99	99	98	98	98
M11 カイ二乗距離	99	99	98	98	98
M12 Rogers and Tanimoto	99	99	98	98	98
M13 Sokal and Sneath 1	99	99	98	98	98
M14 Simple matching	99	99	98	98	98
M15 Hamann	99	99	98	98	98
M16 主効果の相乗平均	99	99	98	98	98

表 A-1-2-1. クラスターの再現性評価率 - ward 法

非類似度	10 分類			15 分類			20 分類			25 分類			30 分類		
	一致度	p 値		一致度	p 値		一致度	p 値		一致度	p 値		一致度	p 値	
M1 交互作用統計量	0.41	0.0%	***	0.35	0.0%	***	0.33	0.0%	***	0.31	0.0%	***	0.30	0.0%	***
M2 Cohen's Kappa	0.37	0.0%	***	0.32	0.0%	***	0.31	0.0%	***	0.30	0.0%	***	0.29	0.0%	***
M3 Psi	0.40	0.0%	***	0.36	0.0%	***	0.33	0.0%	***	0.31	0.0%	***	0.29	0.0%	***
M4 Sokal and Sneath 2	0.37	0.0%	***	0.31	0.0%	***	0.28	0.0%	***	0.27	0.0%	***	0.26	0.0%	***
M5 Jaccard	0.37	0.0%	***	0.31	0.0%	***	0.28	0.0%	***	0.28	0.0%	***	0.27	0.0%	***
M6 Czesanowski	0.37	0.0%	***	0.32	0.0%	***	0.30	0.0%	***	0.28	0.0%	***	0.28	0.0%	***
M7 Kulczynski	0.36	0.0%	***	0.32	0.0%	***	0.28	0.0%	***	0.26	0.0%	***	0.25	0.0%	***
M8 Ochiai	0.36	0.0%	***	0.33	0.0%	***	0.32	0.0%	***	0.32	0.0%	***	0.32	0.0%	***
M9 Yule's Q	0.41	0.0%	***	0.38	0.0%	***	0.38	0.0%	***	0.36	0.0%	***	0.36	0.0%	***
M10 Russel and Rao	0.38	0.0%	***	0.39	0.0%	***	0.35	0.0%	***	0.32	0.0%	***	0.29	0.0%	***
M11 カイ二乗距離	0.34	0.0%	***	0.39	0.0%	***	0.43	0.0%	***	0.44	0.0%	***	0.44	0.0%	***
M12 Rogers and Tanimoto	0.27	0.0%	***	0.26	0.0%	***	0.29	0.0%	***	0.30	0.0%	***	0.29	0.0%	***
M13 Sokal and Sneath 1	0.62	0.0%	***	0.58	0.0%	***	0.59	0.0%	***	0.55	0.0%	***	0.55	0.0%	***
M14 Simple matching	0.57	0.0%	***	0.59	0.0%	***	0.58	0.0%	***	0.56	0.0%	***	0.55	0.0%	***
M15 Hamann	-0.01	12%		-0.02	26%		-0.03	100%		-0.04	100%		-0.02	74%	
M16 主効果の相乗平均	0.59	0.0%	***	0.57	0.0%	***	0.60	0.0%	***	0.59	0.0%	***	0.58	0.0%	***

表 A-1-2-2. クラスターの再現性評価 - 群間平均法

非類似度	10 分類			15 分類			20 分類			25 分類			30 分類		
	一致度	p 値		一致度	p 値	一致度	p 値	一致度	p 値	一致度	p 値		一致度	p 値	
M1 交互作用統計量	0.55	0.0%	***	0.52	0.0%	***	0.50	0.0%	***	0.46	0.0%	***	0.46	0.0%	***
M2 Cohen's Kappa	0.54	0.0%	***	0.49	0.0%	***	0.46	0.0%	***	0.45	0.0%	***	0.45	0.0%	***
M3 Psi	0.50	0.0%	***	0.51	0.0%	***	0.48	0.0%	***	0.45	0.0%	***	0.44	0.0%	***
M4 Sokal and Sneath 2	0.48	0.0%	***	0.46	0.0%	***	0.43	0.0%	***	0.41	0.0%	***	0.39	0.0%	***
M5 Jaccard	0.48	0.0%	***	0.45	0.0%	***	0.43	0.0%	***	0.40	0.0%	***	0.40	0.0%	***
M6 Czesanowski	0.48	0.0%	***	0.45	0.0%	***	0.42	0.0%	***	0.40	0.0%	***	0.40	0.0%	***
M7 Kulczynski	0.49	0.0%	***	0.46	0.0%	***	0.43	0.0%	***	0.41	0.0%	***	0.42	0.0%	***
M8 Ochiai	0.49	0.0%	***	0.49	0.0%	***	0.46	0.0%	***	0.44	0.0%	***	0.41	0.0%	***
M9 Yule's Q	0.46	0.0%	***	0.40	0.0%	***	0.37	0.0%	***	0.34	0.0%	***	0.34	0.0%	***
M10 Russel and Rao	-0.01	94%		0.02	85%		0.04	49%		0.04	50%		0.07	19%	
M11 カイ二乗距離	-0.01	13%		-0.02	27%		-0.03	100%		-0.04	100%		-0.04	100%	
M12 Rogers and Tanimoto	0.50	0.0%	***	0.60	0.0%	***	0.54	0.0%	***	0.54	0.0%	***	0.53	0.0%	***
M13 Sokal and Sneath 1	0.11	14%		0.10	4.2%	*	0.08	12%		0.08	25%		0.11	5.0%	*
M14 Simple matching	0.12	14%		0.08	30%		0.08	12%		0.07	28%		0.10	5.2%	
M15 Hamann	-0.01	11%		-0.02	26%		-0.03	100%		-0.04	100%		-0.05	100%	
M16 主効果の相乗平均	-0.01	12%		-0.02	25%		-0.03	100%		-0.04	100%		-0.05	100%	

表 A-1-2-3. クラスターの再現性評価 - 最長距離法

非類似度	10 分類			15 分類			20 分類			25 分類			30 分類		
	一致度	p 値		一致度	p 値		一致度	p 値		一致度	p 値		一致度	p 値	
M1 交互作用統計量	0.19	0.0%	***	0.15	0.0%	***	0.14	0.0%	***	0.14	0.0%	***	0.15	0.0%	***
M2 Cohen's Kappa	0.19	0.0%	***	0.17	0.0%	***	0.18	0.0%	***	0.19	0.0%	***	0.20	0.0%	***
M3 Psi	-0.18	100%		-0.15	100%		-0.10	100%		-0.06	100%		-0.02	100%	
M4 Sokal and Sneath 2	-0.20	100%		-0.15	100%		-0.10	100%		-0.05	100%		0.00	96%	
M5 Jaccard	-0.19	100%		-0.15	100%		-0.10	100%		-0.04	100%		0.00	95%	
M6 Czesanowski	-0.19	100%		-0.15	100%		-0.10	100%		-0.05	100%		0.00	79%	
M7 Kulczynski	-0.20	100%		-0.15	100%		-0.10	100%		-0.04	100%		0.00	96%	
M8 Ochiai	-0.19	100%		-0.14	100%		-0.09	100%		-0.04	100%		0.00	96%	
M9 Yule's Q	-0.17	100%		-0.12	100%		-0.07	100%		-0.02	100%		0.02	18%	
M10 Russel and Rao	-0.20	100%		-0.17	100%		-0.10	100%		-0.05	100%		0.00	96%	
M11 カイ二乗距離	-0.06	100%		-0.07	100%		-0.09	100%		-0.06	100%		-0.03	100%	
M12 Rogers and Tanimoto	0.06	98%		0.05	85%		0.06	95%		0.12	49%		0.09	25%	
M13 Sokal and Sneath 1	0.21	49%		0.21	19%		0.31	0.1%	**	0.21	0.1%	***	0.19	0.2%	**
M14 Simple matching	0.19	50%		0.22	12%		0.30	0.4%	**	0.24	0.0%	***	0.19	0.1%	***
M15 Hamann	0.05	0.0%	***	0.04	6.5%		0.04	46%		0.04	5.7%		0.04	48%	
M16 主効果の相乗平均	-0.01	12%		-0.02	27%		-0.03	100%		-0.04	100%		-0.05	100%	

表 A-1-2-4. クラスターの再現性評価 - 最短距離法

非類似度	10 分類			15 分類			20 分類			25 分類			30 分類		
	一致度	p 値		一致度	p 値		一致度	p 値		一致度	p 値		一致度	p 値	
M1 交互作用統計量	0.04	100%		0.07	39%		0.07	69%		0.10	26%		0.12	13.9%	
M2 Cohen's Kappa	0.03	100%		0.04	39%		0.05	71%		0.08	32%		0.10	17.8%	
M3 Psi	0.03	100%		0.08	40%		0.08	30%		0.08	26%		0.09	31.0%	
M4 Sokal and Sneath 2	0.05	19%		0.06	40%		0.07	31%		0.10	11%		0.12	5.7%	
M5 Jaccard	0.04	100%		0.05	43%		0.09	28%		0.11	12%		0.13	5.4%	
M6 Czesanowski	0.06	20%		0.07	42%		0.08	28%		0.11	11%		0.12	6.9%	
M7 Kulczynski	0.06	100%		0.07	44%		0.09	31%		0.10	9%		0.14	2.2%	*
M8 Ochiai	0.05	17%		0.05	41%		0.07	29%		0.08	29%		0.12	4.5%	*
M9 Yule's Q	0.00	100%		-0.01	100%		-0.01	100%		0.00	70%		-0.01	82.5%	
M10 Russel and Rao	-0.01	12%		-0.02	26%		-0.03	100%		-0.03	100%		-0.04	100.0%	
M11 カイ二乗距離	-0.01	12%		-0.02	26%		-0.03	100%		-0.04	100%		-0.05	100.0%	
M12 Rogers and Tanimoto	-0.01	12%		-0.01	100%		-0.01	100%		-0.02	100%		-0.01	73.0%	
M13 Sokal and Sneath 1	-0.01	100%		-0.01	100%		-0.01	100%		-0.01	100%		-0.01	74.1%	
M14 Simple matching	-0.01	100%		-0.02	100%		-0.01	100%		-0.01	100%		0.00	74.3%	
M15 Hamann	-0.01	12%		-0.02	26%		-0.03	100%		-0.04	100%		-0.05	100.0%	
M16 主効果の相乗平均	-0.01	11%		-0.02	26%		-0.03	100%		-0.04	100%		-0.05	100.0%	

表 A-1-3-1. 外的基準との独立性検証 - 製品分類数別 PAM 法

非類似度	χ^2 統計量の順位					$\chi^2/d.f.$				
	10 分類	15 分類	20 分類	25 分類	30 分類	10 分類	15 分類	20 分類	25 分類	30 分類
M1 交互作用統計量	1	4	8	1	1	9.1	7.1	5.9	6.1	5.8
M2 Cohen's Kappa	6.5	5	2	5	4.5	7.9	7.0	6.6	5.7	5.1
M3 Psi	6.5	8	9	3	2	7.9	6.7	5.8	6.0	5.6
M4 Sokal and Sneath 2	3.5	6.5	7	7	8	8.2	6.8	5.9	5.0	4.7
M5 Jaccard	3.5	6.5	5	8	9	8.2	6.8	6.0	5.0	4.7
M6 Czesanowski	10	1	5	6	4.5	6.9	7.3	6.0	5.6	5.1
M7 Kulczynski	8	9	10	10	10	7.1	6.4	5.7	4.7	4.5
M8 Ochiai	2	2	3	2	3	8.8	7.3	6.6	6.1	5.3
M9 Yule's Q	9	10	5	9	7	6.9	5.7	6.0	4.7	5.0
M10 Russel and Rao	5	3	1	4	6	8.0	7.2	6.8	5.8	5.0
M11 カイ二乗距離	15	15	11	11	11	1.3	1.1	2.3	1.9	1.9
M12 Rogers and Tanimoto	11	11	12	12	12	3.6	2.5	1.9	1.6	1.6
M13 Sokal and Sneath 1	12.5	12.5	13.5	13.5	13.5	3.3	2.3	1.7	1.4	1.2
M14 Simple matching	12.5	12.5	13.5	13.5	13.5	3.3	2.3	1.7	1.4	1.2
M15 Hamann	16	16	16	16	16	0.4	0.3	0.2	0.2	0.2
M16 主効果の相乗平均	14	14	15	15	15	2.7	2.1	1.6	1.3	1.2

表 A-1-3-2. 外的基準との独立性検証 - 製品分類数別 ward 法

非類似度	χ^2 統計量の順位					$\chi^2/d.f.$				
	10 分類	15 分類	20 分類	25 分類	30 分類	10 分類	15 分類	20 分類	25 分類	30 分類
M1 交互作用統計量	5	2	3	2	3	10.0	8.7	7.3	6.4	5.8
M2 Cohen's Kappa	4	4	5	5	5	10.0	8.5	7.1	6.2	5.4
M3 Psi	2	3	4	3.5	4	10.2	8.5	7.3	6.3	5.6
M4 Sokal and Sneath 2	8	8	8	8	8	8.9	7.3	6.4	5.6	5.0
M5 Jaccard	7	7	7	7	7	8.9	7.6	6.5	5.8	5.1
M6 Czesanowski	6	6	6	6	6	9.3	7.9	6.6	5.9	5.2
M7 Kulczynski	9	9	9	9	9	8.6	7.1	6.3	5.5	5.0
M8 Ochiai	3	1	2	1	1	10.2	8.7	7.4	6.6	5.8
M9 Yule's Q	1	5	1	3.5	2	10.2	8.5	7.4	6.3	5.8
M10 Russel and Rao	11	10	10	10	10	5.8	5.2	4.8	4.5	4.2
M11 カイ二乗距離	10	11	11	11	11	6.0	5.0	4.4	4.1	3.7
M12 Rogers and Tanimoto	13	12	12	13	13	1.6	1.6	1.4	1.4	1.4
M13 Sokal and Sneath 1	16	16	15	15	16	1.1	1.1	1.1	1.1	1.3
M14 Simple matching	15	14	14	14	15	1.3	1.2	1.1	1.1	1.3
M15 Hamann	14	15	16	16	14	1.5	1.2	1.0	1.1	1.3
M16 主効果の相乗平均	12	13	13	12	12	2.0	1.5	1.2	1.5	1.5

表 A-1-3-3. 外的基準との独立性検証 - 製品分類数別 群間平均法

非類似度	χ^2 統計量の順位					$\chi^2/d.f.$				
	10 分類	15 分類	20 分類	25 分類	30 分類	10 分類	15 分類	20 分類	25 分類	30 分類
M1 交互作用統計量	1	1	1	1	1	9.5	7.6	6.7	6.1	5.7
M2 Cohen's Kappa	2	3	3	4	4	9.1	7.3	6.4	5.8	5.3
M3 Psi	5	2	2	2	2	8.8	7.4	6.5	6.0	5.6
M4 Sokal and Sneath 2	9	8	6	8	7	8.5	7.0	6.2	5.5	5.2
M5 Jaccard	7	6	7	7	5	8.6	7.0	6.1	5.6	5.3
M6 Czesanowski	6	9	8	5	8	8.6	6.9	6.1	5.7	5.2
M7 Kulczynski	8	5	5	6	6	8.6	7.0	6.2	5.6	5.2
M8 Ochiai	3	4	4	3	3	8.9	7.1	6.3	5.8	5.4
M9 Yule's Q	4	7	9	9	9	8.8	7.0	6.0	5.2	4.9
M10 Russel and Rao	10	10	10	10	10	4.1	4.0	3.9	3.9	3.8
M11 カイ二乗距離	11	11	11	11	11	1.9	1.8	1.8	2.2	2.0
M12 Rogers and Tanimoto	14	14	14	16	16	0.3	0.3	0.3	0.4	0.8
M13 Sokal and Sneath 1	16	15.5	15	14	15	0.2	0.2	0.3	0.5	0.8
M14 Simple matching	15	15.5	16	15	14	0.2	0.2	0.2	0.5	0.9
M15 Hamann	13	13	13	13	13	1.5	1.2	1.0	1.0	1.4
M16 主効果の相乗平均	12	12	12	12	12	1.7	1.4	1.7	1.6	1.5

表 A-1-3-4. 外的基準との独立性検証 - 製品分類数別 最長距離法

非類似度	χ^2 統計量の順位					$\chi^2/d.f.$				
	10 分類	15 分類	20 分類	25 分類	30 分類	10 分類	15 分類	20 分類	25 分類	30 分類
M1 交互作用統計量	2	2	2	2	2	7.3	5.6	4.8	4.3	4.1
M2 Cohen's Kappa	1	1	1	1	1	7.5	5.7	5.1	4.7	4.3
M3 Psi	10	10	10	10	6	3.6	3.7	3.6	3.8	3.9
M4 Sokal and Sneath 2	9	6.5	8	9	7	3.9	4.0	3.8	3.9	3.9
M5 Jaccard	7	9	5.5	5	8	4.0	3.9	3.9	4.0	3.8
M6 Czesanowski	4	5	5.5	7	9	4.2	4.0	3.9	3.9	3.8
M7 Kulczynski	6	4	7	6	10	4.0	4.1	3.9	4.0	3.7
M8 Ochiai	5	8	4	4	4	4.0	4.0	4.0	4.0	4.0
M9 Yule's Q	3	3	3	3	3	4.3	4.2	4.2	4.1	4.1
M10 Russel and Rao	8	6.5	9	8	5	3.9	4.0	3.8	3.9	3.9
M11 カイ二乗距離	11	11	11	11	11	2.4	2.5	2.5	2.4	2.5
M12 Rogers and Tanimoto	16	16	14	16	16	1.0	0.9	0.7	0.7	0.9
M13 Sokal and Sneath 1	15	15	15	14	15	1.2	0.9	0.7	0.8	1.2
M14 Simple matching	14	14	16	15	14	1.3	0.9	0.6	0.8	1.2
M15 Hamann	13	12	13	13	13	1.4	1.6	1.4	1.3	1.4
M16 主効果の相乗平均	12	13	12	12	12	1.7	1.5	1.7	1.6	1.5

表 A-1-3-5. 外的基準との独立性検証 - 製品分類数別 最短距離法

非類似度	χ^2 統計量の順位					$\chi^2/d.f.$				
	10 分類	15 分類	20 分類	25 分類	30 分類	10 分類	15 分類	20 分類	25 分類	30 分類
M1 交互作用統計量	9	9	4.5	1	2	3.1	3.5	3.6	3.6	3.5
M2 Cohen's Kappa	7	8	2	2	3	3.5	3.8	3.9	3.6	3.3
M3 Psi	8	5.5	3	4	1	3.2	3.8	3.7	3.5	3.7
M4 Sokal and Sneath 2	3	5.5	6	7	6	4.4	3.8	3.5	3.2	3.1
M5 Jaccard	2	7	8	6	7	4.5	3.8	3.3	3.2	3.0
M6 Czesanowski	6	3	9	8	9	4.1	4.0	3.2	3.2	3.0
M7 Kulczynski	1	4	7	9	5	4.6	3.9	3.4	3.2	3.1
M8 Ochiai	5	2	4.5	5	8	4.2	4.1	3.6	3.4	3.0
M9 Yule's Q	11	12	12	12	12	1.7	1.2	1.0	1.1	1.0
M10 Russel and Rao	4	1	1	3	4	4.4	4.5	4.0	3.5	3.3
M11 カイ二乗距離	10	10	10	10	10	1.8	1.8	2.0	2.2	2.0
M12 Rogers and Tanimoto	16	15	15	15	15	0.2	0.2	0.3	0.5	0.8
M13 Sokal and Sneath 1	15	15	16	14	14	0.2	0.2	0.3	0.6	0.9
M14 Simple matching	14	15	13	13	13	0.2	0.2	0.3	0.6	0.9
M15 Hamann	13	13	14	16	16	0.3	0.4	0.3	0.3	0.3
M16 主効果の相乗平均	12	11	11	11	11	1.7	1.6	1.7	1.6	1.5

表 A-1-4-1. 消費者分類 × 製品分類の度数表における独立性検証 PAM 法

非類似度			製品分類									
			χ^2 順位					$\chi^2/d.f.$				
			10 分類	15 分類	20 分類	25 分類	30 分類	10 分類	15 分類	20 分類	25 分類	30 分類
消費者分類	M1 交互作用統計量	10 分類	2	2	1	2	2	715	486	373	303	254
		15 分類	1	1	1	1	1	499	350	270	218	184
		20 分類	1	1	1	1	1	378	271	210	170	144
		25 分類	1	1	2	2	2	315	229	179	146	123
		30 分類	1	1	1	1	1	265	195	153	125	106
	M2 Cohen's Kappa	10 分類	3	3	3	3	3	672	475	365	296	247
		15 分類	5	4	4	4	5	449	323	250	204	170
		20 分類	6	4	3.5	4	6	339	248	193	159	132
		25 分類	5	4	5	3	4	285	211	164	138	115
		30 分類	6	4	4	4	4	242	182	141	118	99
	M3 Psi	10 分類	8	7	4	4	4	620	451	359	292	245
		15 分類	6	3	2	2	2	448	330	265	215	180
		20 分類	5	2	2	2	2	347	259	208	169	142
		25 分類	4	2	1	1	1	289	221	181	148	125
		30 分類	5	3	2	2	2	242	186	151	124	105
	M4 Sokal and Sneath 2	10 分類	4.5	4.5	5	7	5	667	466	355	282	244
		15 分類	3.5	5.5	5	6	4	451	317	245	195	171
		20 分類	3.5	5.5	6	7	4.5	348	246	192	153	133
		25 分類	3	5	3	6	5	294	211	166	132	115
		30 分類	3	6	6	7	6	246	177	139	112	97
	M5 Jaccard	10 分類	4.5	4.5	7.5	6	7	667	466	353	283	238
		15 分類	3.5	5.5	6	5	7	451	317	245	196	165
		20 分類	3.5	5.5	7	6	7	348	246	192	154	130
		25 分類	6	7	7	7	7	279	200	156	126	108
		30 分類	4	7	7	6	7	244	176	138	112	95
	M6 Czesanowski	10 分類	7	8	7.5	5	6	621	447	353	285	243
		15 分類	8	7	7	7	6	408	305	241	194	166
		20 分類	8	7	3.5	5	4.5	326	244	193	155	133
		25 分類	7	6	4	5	6	274	208	164	133	115
		30 分類	8	5	5	5	5	230	179	141	115	99
	M7 Kulczynski	10 分類	9	6	6	8	8	600	453	354	277	235
		15 分類	9	8	8	8	8	403	302	237	186	159
		20 分類	9	8	8	8	8	309	233	188	149	128
		25 分類	9	8	8	8	8	254	191	155	123	105
		30 分類	9	8	8	8	8	223	170	137	110	94
	M8 Ochiai	10 分類	1	1	2	1	1	723	495	372	304	255
		15 分類	2	2	3	3	3	484	337	252	207	174
		20 分類	2	3	5	3	3	365	257	193	159	135
		25 分類	2	3	6	4	3	299	218	164	137	116
		30 分類	2	2	3	3	3	259	190	143	120	102

表 A-1-4-1. 消費者分類 × 製品分類の度数表における独立性検証 PAM 法 (続き)

非類似度			製品分類									
			χ^2 順位					$\chi^2/d.f.$				
			10 分類	15 分類	20 分類	25 分類	30 分類	10 分類	15 分類	20 分類	25 分類	30 分類
消費者分類	M9 Yule's Q	10 分類	10	10	10	10	10	400	335	276	227	186
		15 分類	10	10	10	10	10	272	236	199	161	135
		20 分類	10	10	9	9	9	228	213	183	147	124
		25 分類	10	10	9	9	9	185	172	148	119	102
		30 分類	10	10	9	9	9	156	146	125	102	87
	M10 Russel and Rao	10 分類	6	9	9	9	9	638	425	317	252	209
		15 分類	7	9	9	9	9	433	290	216	172	143
		20 分類	7	9	10	10	10	329	221	166	132	110
		25 分類	8	9	10	10	10	272	183	137	109	91
		30 分類	7	9	10	10	10	230	155	116	93	77
	M11 カイ二乗距離	10 分類	12	14	14	14	14	224	145	111	89	75
		15 分類	14	14	14	14	14	145	94	72	58	49
		20 分類	12	14	14	14	14	127	82	63	53	45
		25 分類	12	14	14	14	14	99	64	50	42	35
		30 分類	12	14	14	14	14	95	63	49	41	34
	M12 Rogers and Tanimoto	10 分類	11	11	11	11	11	376	264	201	168	150
		15 分類	11	11	11	11	11	252	178	137	116	104
		20 分類	11	11	11	11	11	194	138	106	90	81
		25 分類	11	11	11	11	11	156	112	86	73	66
		30 分類	11	11	11	11	11	139	99	76	64	59
	M13 Sokal and Sneath 1	10 分類	13.5	12.5	12.5	12.5	12.5	224	237	186	155	136
		15 分類	12.5	12.5	12.5	12.5	12.5	155	174	136	113	100
		20 分類	13.5	12.5	12.5	12.5	12.5	117	132	103	87	77
		25 分類	13.5	12.5	12.5	12.5	12.5	98	109	85	71	62
		30 分類	13.5	12.5	12.5	12.5	12.5	84	92	73	61	53
	M14 Simple matching	10 分類	13.5	12.5	12.5	12.5	12.5	224	237	186	155	136
		15 分類	12.5	12.5	12.5	12.5	12.5	155	174	136	113	100
		20 分類	13.5	12.5	12.5	12.5	12.5	117	132	103	87	77
		25 分類	13.5	12.5	12.5	12.5	12.5	98	109	85	71	62
		30 分類	13.5	12.5	12.5	12.5	12.5	84	92	73	61	53
	M15 Hamann	10 分類	15	15	15	15	15	9	6	5	4	3
		15 分類	15	15	15	15	15	6	4	3	3	2
		20 分類	15	15	15	15	15	4	3	2	2	2
		25 分類	15	15	15	15	15	3	2	2	2	1
		30 分類	15	15	15	15	15	3	2	2	1	1
	M16 主効果の相乗平均	10 分類	16	16	16	16	16	0	0	0	0	0
		15 分類	16	16	16	16	16	0	0	0	0	0
		20 分類	16	16	16	16	16	2	1	1	1	1
		25 分類	16	16	16	16	16	1	1	1	1	0
		30 分類	16	16	16	16	16	1	1	1	0	0

表 A-1-4-2. 消費者分類 × 製品分類の度数表における独立性検証 ward 法

非類似度			製品分類									
			χ^2 順位					$\chi^2/d.f.$				
			10 分類	15 分類	20 分類	25 分類	30 分類	10 分類	15 分類	20 分類	25 分類	30 分類
消費者分類	M1 交互作用統計量	10 分類	3	3	3	3	3	794	546	415	334	280
		15 分類	2	1	2	2	2	555	388	297	241	202
		20 分類	1	1	1	1	1	430	305	235	191	161
		25 分類	2	1	1	1	1	350	250	194	158	133
		30 分類	1	1	1	1	1	296	212	165	135	114
	M2 Cohen's Kappa	10 分類	1	1	1	1	1	805	550	419	337	283
		15 分類	3	3	3	4	3.5	554	385	296	239	202
		20 分類	3	4	3	4	4	427	300	233	189	159
		25 分類	4	4	4	4	4	347	247	191	156	132
		30 分類	3	4	4	4	4	293	209	163	133	113
	M3 Psi	10 分類	4	4	4	4	4	779	537	410	331	278
		15 分類	4	4	4	3	3.5	548	385	295	240	202
		20 分類	4	3	4	3	3	424	301	233	189	160
		25 分類	3	2	3	3	2	347	248	192	157	133
		30 分類	4	3	3	3	3	292	210	163	134	113
	M4 Sokal and Sneath 2	10 分類	7	7	7	7	6	747	523	403	327	276
		15 分類	6	6	5	5	5	525	372	288	235	199
		20 分類	6	7	5	6	5	404	289	225	184	156
		25 分類	7	7	6	6	5	329	237	185	152	129
		30 分類	6	7	6	6	5	278	200	157	129	110
	M5 Jaccard	10 分類	6	6	6	5	5	751	527	403	329	276
		15 分類	7	7	7	6	6	521	370	286	234	197
		20 分類	7	6	6	5	6	403	289	225	184	156
		25 分類	6	6	5	5	6	329	237	186	152	129
		30 分類	7	5	5	5	6	278	202	158	130	110
	M6 Czesanowski	10 分類	5	5	5	6	7	770	532	405	328	276
		15 分類	5	5	6	7	7	529	373	286	233	197
		20 分類	5	5	7	8	8	409	290	223	182	154
		25 分類	5	5	7	7	7	334	239	185	151	128
		30 分類	5	6	7	8	7	280	202	157	129	110
	M7 Kulczynski	10 分類	8	8	8	8	8	732	514	398	322	272
		15 分類	8	8	8	8	8	513	366	284	232	196
		20 分類	8	8	8	7	7	398	285	223	182	154
		25 分類	8	8	8	8	8	325	234	185	151	128
		30 分類	8	8	8	7	8	275	199	157	129	110
	M8 Ochiai	10 分類	2	2	2	2	2	804	549	418	337	282
		15 分類	1	2	1	1	1	557	388	298	242	203
		20 分類	2	2	2	2	2	428	302	234	190	160
		25 分類	1	3	2	2	3	350	248	193	157	133
		30 分類	2	2	2	2	2	295	211	164	134	114

表 A-1-4-2. 消費者分類 × 製品分類の度数表における独立性検証 ward 法 (続き)

非類似度			製品分類									
			χ^2 順位					$\chi^2/d.f.$				
			10 分類	15 分類	20 分類	25 分類	30 分類	10 分類	15 分類	20 分類	25 分類	30 分類
消費者分類	M9 Yule's Q	10 分類	9	9	9	9	10	685	496	385	310	260
		15 分類	9	9	9	10	10	464	342	268	217	184
		20 分類	9	9	10	10	10	351	260	205	167	141
		25 分類	9	10	10	10	10	283	210	166	136	115
		30 分類	9	10	10	10	10	237	177	141	115	98
	M10 Russel and Rao	10 分類	10	10	10	10	9	628	480	378	309	262
		15 分類	10	10	10	9	9	427	335	267	220	187
		20 分類	10	10	9	9	9	326	259	209	173	148
		25 分類	10	9	9	9	9	265	212	172	143	123
		30 分類	10	9	9	9	9	223	179	146	122	105
	M11 カイ二乗距離	10 分類	11	11	11	11	11	486	339	264	214	176
		15 分類	11	11	11	11	11	332	238	187	153	126
		20 分類	11	11	11	11	11	255	186	147	122	102
		25 分類	11	11	11	11	11	208	154	122	102	86
		30 分類	11	11	11	11	11	175	131	105	89	75
	M12 Rogers and Tanimoto	10 分類	12	12	12	12	13	246	212	187	171	151
		15 分類	12	12	12	12	12	172	149	131	120	108
		20 分類	12	12	12	12	12	133	116	102	94	85
		25 分類	12	12	12	12	12	109	95	84	77	70
		30 分類	12	12	12	12	12	92	81	72	66	60
	M13 Sokal and Sneath 1	10 分類	14	14	14	14	14	211	186	175	157	147
		15 分類	14	14	14	14	14	145	129	121	109	102
		20 分類	14	14	14	14	14	111	99	94	84	80
		25 分類	14	14	14	14	14	91	81	77	69	65
		30 分類	14	14	14	14	14	77	69	65	59	56
	M14 Simple matching	10 分類	13	13	13	13	12	229	205	186	160	153
		15 分類	13	13	13	13	13	157	141	128	111	107
		20 分類	13	13	13	13	13	121	109	99	86	83
		25 分類	13	13	13	13	13	98	89	81	71	68
		30 分類	13	13	13	13	13	83	75	69	60	58
	M15 Hammann	10 分類	16	16	16	16	16	1	1	1	1	1
		15 分類	16	16	16	16	16	1	1	1	1	1
		20 分類	16	16	16	16	16	1	1	1	1	1
		25 分類	16	16	16	16	16	1	1	1	1	1
		30 分類	16	16	16	16	16	1	1	1	1	1
	M16 主効果の相乗平均	10 分類	15	15	15	15	15	6	4	4	3	3
		15 分類	15	15	15	15	15	4	3	3	2	2
		20 分類	15	15	15	15	15	3	3	2	2	2
		25 分類	15	15	15	15	15	3	2	2	2	2
		30 分類	15	15	15	15	15	3	2	2	2	1

表 A-1-4-3. 消費者分類 × 製品分類の度数表における独立性検証 群間平均法

非類似度			製品分類									
			χ^2 順位					$\chi^2/d.f.$				
			10 分類	15 分類	20 分類	25 分類	30 分類	10 分類	15 分類	20 分類	25 分類	30 分類
消費者分類	M1 Cross	10 分類	3	3	3	6	4	398	275	212	171	147
		15 分類	3	3	3	6	6	275	192	149	121	104
		20 分類	3	4	4	7	8	211	149	118	96	83
		25 分類	3	4	4	7	8	175	123	98	81	70
		30 分類	3	3	3	7	8	151	108	86	71	62
	M2 Cohen	10 分類	1	2	2	2	2	426	292	227	185	158
		15 分類	2	2	2	2	2	293	203	161	133	115
		20 分類	2	2	2	2	2	232	165	131	109	94
		25 分類	1	2	2	2	2	195	138	111	93	81
		30 分類	1	2	2	2	2	167	119	97	81	71
	M3 Psi	10 分類	9	9	9	9	9	349	255	201	164	140
		15 分類	9	9	9	9	9	245	181	144	118	101
		20 分類	9	9	9	9	9	187	139	112	93	80
		25 分類	9	9	8	9	9	157	118	96	80	69
		30 分類	9	9	9	9	9	134	102	82	69	60
	M4 Sokal and Sneath 2	10 分類	7	8	8	8	8	384	266	205	169	143
		15 分類	6	6	7	5	8	263	186	146	122	104
		20 分類	6	5	5	5	5	202	145	115	97	84
		25 分類	6	6	6	5	4	167	120	96	82	72
		30 分類	6	5	6	5	5	144	104	83	72	63
	M5 Jaccard	10 分類	8	6	7	5	5	383	269	209	172	146
		15 分類	8	8	8	7	7	259	184	144	121	104
		20 分類	8	7	8	6	7	200	144	114	97	83
		25 分類	8	7	7	6	6	165	119	96	81	72
		30 分類	8	8	7	6	6	141	103	83	72	63
	M6 Czesanowski	10 分類	5	7	6	4	7	388	267	209	172	145
		15 分類	4	7	6	3	5	266	185	147	123	105
		20 分類	7	8	7	4	4	202	143	115	98	85
		25 分類	5	5	5	4	5	168	120	97	83	72
		30 分類	5	6	5	4	3	145	104	85	74	64
	M7 Kulczynski	10 分類	4	5	5	7	6	390	270	210	171	146
		15 分類	5	5	4	8	4	266	186	147	120	105
		20 分類	5	6	6	8	6	204	144	115	96	84
		25 分類	7	8	9	8	7	166	119	95	80	70
		30 分類	7	7	8	8	7	143	103	83	71	63
	M8 Ochiai	10 分類	6	4	4	3	3	385	274	211	174	149
		15 分類	7	4	5	4	3	262	188	147	123	106
		20 分類	4	3	3	3	3	208	151	120	101	88
		25 分類	4	3	3	3	3	170	124	99	84	73
		30 分類	4	4	4	3	4	145	107	86	74	64

表 A-1-4-3. 消費者分類 × 製品分類の度数表における独立性検証 群間平均法 (続き)

非類似度			製品分類									
			χ^2 順位					$\chi^2/d.f.$				
			10 分類	15 分類	20 分類	25 分類	30 分類	10 分類	15 分類	20 分類	25 分類	30 分類
消費者分類	M9 Yules Q	10 分類	2	1	1	1	1	424	298	233	192	166
		15 分類	1	1	1	1	1	299	214	170	143	125
		20 分類	1	1	1	1	1	237	171	139	117	103
		25 分類	2	1	1	1	1	192	140	114	97	86
		30 分類	2	1	1	1	1	163	119	97	83	74
	M10 Russel and Rao	10 分類	10	10	10	10	10	47	46	44	43	45
		15 分類	10	10	10	10	10	41	40	39	39	39
		20 分類	10	10	10	10	10	34	35	33	32	33
		25 分類	10	10	10	10	10	30	30	28	28	28
		30 分類	10	10	10	10	10	26	26	25	24	25
	M11 Chi	10 分類	11	11	11	11	11	11	10	10	11	10
		15 分類	11	11	11	11	11	12	10	10	10	9
		20 分類	11	11	11	11	11	11	10	9	9	8
		25 分類	11	11	11	11	11	10	9	9	9	8
		30 分類	11	11	11	11	11	10	8	8	8	8
	M12 Rogers and Tanimoto	10 分類	12	12	12	12	12	1	1	1	1	1
		15 分類	12	12	12	12	12	1	1	1	1	1
		20 分類	12	12	12	12	12	1	1	1	1	1
		25 分類	12	12	12	12	12	1	1	1	1	1
		30 分類	12	12	12	12	12	1	1	1	1	1
	M13 Sokal and Sneath 1	10 分類	15.5	14	13	15	13	1	1	1	1	1
		15 分類	15	15	15	15	14	1	1	1	1	1
		20 分類	16	14	15	14	15	1	1	1	1	1
		25 分類	15	14	15	14	14	1	1	1	1	1
		30 分類	13	13	13	13	14	1	1	1	1	1
	M14 Simple	10 分類	15.5	15	14	16	14	1	1	1	1	1
		15 分類	14	14	14	14	13	1	1	1	1	1
		20 分類	15	13	14	15	14	1	1	1	1	1
		25 分類	14	13	14	13	13	1	1	1	1	1
		30 分類	14	14	14	14	13	1	1	1	1	1
	M15 Hammann	10 分類	13	13	16	14	16	1	1	1	1	1
		15 分類	13	13	16	16	16	1	1	1	0	0
		20 分類	13	15	16	16	16	1	1	0	0	0
		25 分類	16	16	16	16	16	1	0	0	0	0
		30 分類	16	16	16	16	16	1	0	0	0	0
	M16 Volume	10 分類	14	16	15	13	15	1	0	1	1	1
		15 分類	16	16	13	13	15	1	1	1	1	1
		20 分類	14	16	13	13	13	1	0	1	1	1
		25 分類	13	15	13	15	15	1	1	1	1	1
		30 分類	15	15	15	15	15	1	1	1	1	1

表 A-1-4-4. 消費者分類 × 製品分類の度数表における独立性検証 最長距離法

非類似度			製品分類									
			χ^2 順位					$\chi^2/d.f.$				
			10 分類	15 分類	20 分類	25 分類	30 分類	10 分類	15 分類	20 分類	25 分類	30 分類
消費者分類	M1 Cross	10 分類	2	2	2	2	2	291	203	158	132	113
		15 分類	2	2	2	2	2	203	144	115	97	83
		20 分類	2	2	2	2	2	154	111	89	76	66
		25 分類	2	2	2	2	2	127	92	75	64	55
		30 分類	2	2	2	2	2	109	80	65	57	49
	M2 Cohen	10 分類	1	1	1	1	1	328	229	183	152	130
		15 分類	1	1	1	1	1	243	174	142	120	103
		20 分類	1	1	1	1	1	196	142	117	100	87
		25 分類	1	1	1	1	1	163	121	100	86	75
		30 分類	1	1	1	1	1	141	104	87	75	65
	M3 Psi	10 分類	14	12	10	10	10	20	24	26	27	26
		15 分類	14	11	10	10	10	19	23	25	26	25
		20 分類	13	11	10	10	10	18	21	23	24	23
		25 分類	13	11	10	10	10	17	19	22	22	22
		30 分類	13	11	10	10	10	16	18	20	21	21
	M4 Sokal and Sneath 2	10 分類	9	7	7	8	5	38	39	38	36	35
		15 分類	7	6	5	8	7	34	33	33	32	30
		20 分類	7	8	7	8	7	30	30	29	28	27
		25 分類	7	8	8	8	8	28	27	27	26	24
		30 分類	7	8	6	8	6	25	25	25	24	23
	M5 Jaccard	10 分類	5	5	6	4	8	43	39	38	38	34
		15 分類	6	5	8	6	6	36	34	32	32	30
		20 分類	6	5	5	5	5	32	31	29	29	27
		25 分類	5	5	6	6	6	29	28	27	27	25
		30 分類	6	6	8	6	7	26	25	25	24	23
	M6 Czesanowski	10 分類	4	6	5	5	7	44	39	39	37	34
		15 分類	4	7	7	7	5	37	33	32	32	30
		20 分類	4	7	8	7	6	33	30	29	29	27
		25 分類	4	6	5	7	5	30	28	27	27	25
		30 分類	4	7	7	7	8	27	25	25	24	22
	M7 Kulczynski	10 分類	8	8	8	7	9	38	39	38	36	34
		15 分類	8	8	6	5	8	33	33	33	32	29
		20 分類	8	6	6	6	8	30	31	29	29	27
		25 分類	8	7	7	5	7	27	28	27	27	25
		30 分類	8	4	4	5	5	25	26	25	25	23
	M8 Ochiai	10 分類	6	4	4	6	4	42	40	39	37	35
		15 分類	5	4	4	4	4	37	35	34	33	31
		20 分類	5	4	4	4	4	32	32	30	30	28
		25 分類	6	4	4	4	4	29	28	28	27	26
		30 分類	5	5	5	4	4	27	26	25	25	24

表 A-1-4-4. 消費者分類 × 製品分類の度数表における独立性検証 最長距離法 (続き)

非類似度			製品分類									
			χ^2 順位					$\chi^2/d.f.$				
			10 分類	15 分類	20 分類	25 分類	30 分類	10 分類	15 分類	20 分類	25 分類	30 分類
消費者分類	M9 Yules Q	10 分類	3	3	3	3	3	46	46	42	40	37
		15 分類	3	3	3	3	3	41	39	37	35	32
		20 分類	3	3	3	3	3	36	35	34	32	30
		25 分類	3	3	3	3	3	32	32	31	29	27
		30 分類	3	3	3	3	3	30	30	29	27	26
	M10 Russel and Rao	10 分類	7	9	9	9	6	39	37	37	36	34
		15 分類	9	9	9	9	9	33	30	30	30	29
		20 分類	9	9	9	9	9	28	27	28	27	26
		25 分類	9	9	9	9	9	25	24	25	24	24
		30 分類	9	9	9	9	9	24	22	23	23	22
	M11 Chi	10 分類	12	10	11	11	11	30	29	25	25	24
		15 分類	10	10	11	11	11	26	25	22	22	21
		20 分類	10	10	11	11	11	25	24	21	20	19
		25 分類	10	10	11	11	11	22	22	20	19	18
		30 分類	10	10	11	11	11	21	21	19	18	18
	M12 Rogers and Tanimoto	10 分類	13	14	14	14	14	21	17	15	14	14
		15 分類	13	14	14	14	14	20	16	14	13	13
		20 分類	14	12	12	13	14	18	15	13	12	12
		25 分類	14	13	12	14	14	17	13	12	11	12
		30 分類	14	12	12	14	14	16	13	12	11	11
	M13 Sokal and Sneath 1	10 分類	11	13	12	13	12	32	23	20	19	20
		15 分類	11	12	12	12	12	24	18	16	15	16
		20 分類	12	14	13	14	12	19	14	13	12	13
		25 分類	11	12	13	12	12	18	14	12	12	12
		30 分類	12	14	13	12	12	17	12	11	11	11
	M14 Simple	10 分類	10	11	13	12	13	34	24	20	19	20
		15 分類	12	13	13	13	13	23	17	15	15	14
		20 分類	11	13	14	12	13	20	14	13	13	13
		25 分類	12	14	14	13	13	17	13	11	11	12
		30 分類	11	13	14	13	13	17	13	11	11	11
	M15 Hammann	10 分類	15	15	15	15	15	2	2	1	1	1
		15 分類	15	15	15	15	15	2	2	1	1	1
		20 分類	15	15	15	15	15	2	2	2	2	1
		25 分類	15	15	15	15	15	2	2	2	2	2
		30 分類	15	15	15	15	15	2	2	2	2	2
	M16 Volume	10 分類	16	16	16	16	16	0	0	1	1	1
		15 分類	16	16	16	16	16	1	1	1	1	1
		20 分類	16	16	16	16	16	1	1	1	1	1
		25 分類	16	16	16	16	16	1	1	1	1	1
		30 分類	16	16	16	16	16	1	1	1	1	1

表 A-1-4-5. 消費者分類 × 製品分類の度数表における独立性検証 最短距離法

非類似度			製品分類									
			χ^2 順位					$\chi^2/d.f.$				
			10 分類	15 分類	20 分類	25 分類	30 分類	10 分類	15 分類	20 分類	25 分類	30 分類
消費者分類	M1 Cross	10 分類	8	10	7	6	6	3	2	3	3	3
		15 分類	6	4	4	4	4	3	3	3	3	3
		20 分類	5	5	4	5	7	2	2	3	3	2
		25 分類	3	4	6	5	5	3	3	2	3	2
		30 分類	3	4	6	5	4	3	2	2	3	3
	M2 Cohen	10 分類	10	8	10	10	10	2	2	2	2	2
		15 分類	10	10	10	10	10	2	2	2	2	2
		20 分類	10	10	10	10	10	2	2	2	2	2
		25 分類	10	10	10	10	10	2	2	2	2	2
		30 分類	10	10	10	10	10	2	2	2	2	2
	M3 Psi	10 分類	9	9	8	7	9	2	2	3	3	3
		15 分類	9	5	7	5	8	2	3	3	3	2
		20 分類	6	4	6	4	4	2	3	2	3	3
		25 分類	4	5	4	4	4	3	3	3	3	3
		30 分類	5	3	5	3	3	2	3	2	3	3
	M4 Sokal and Sneath 2	10 分類	7	7	9	9	4	3	3	3	3	3
		15 分類	5	7	6	8	5	3	2	3	2	3
		20 分類	8	9	7	8	6	2	2	2	2	2
		25 分類	8	6	8	8	9	2	2	2	2	2
		30 分類	9	7	9	6	8	2	2	2	2	2
	M5 Jaccard	10 分類	6	5	6	8	5	3	3	3	3	3
		15 分類	8	8	9	6	7	2	2	2	2	2
		20 分類	7	7	9	7	9	2	2	2	2	2
		25 分類	5	8	7	7	8	3	2	2	2	2
		30 分類	4	9	7.5	8	6	2	2	2	2	2
	M6 Czesanowski	10 分類	4	6	4	4	8	3	3	3	3	3
		15 分類	7	6	5	7	9	3	2	3	2	2
		20 分類	9	6	8	9	5	2	2	2	2	3
		25 分類	7	9	9	9	7	2	2	2	2	2
		30 分類	7	8	7.5	7	9	2	2	2	2	2
	M7 Kulczynski	10 分類	3	4	5	5	7	3	3	3	3	3
		15 分類	4	9	8	9	6	3	2	2	2	3
		20 分類	4	8	5	6	8	2	2	2	2	2
		25 分類	9	7	3	6	6	2	2	3	2	2
		30 分類	8	6	3	9	7	2	2	2	2	2
	M8 Ochiai	10 分類	5	3	3	3	3	3	3	3	3	3
		15 分類	3	3	3	3	3	3	3	3	3	3
		20 分類	3	3	3	3	3	2	3	3	3	3
		25 分類	2	3	5	3	3	3	3	3	3	3
		30 分類	6	5	4	4	5	2	2	2	3	2

表 A-1-4-5. 消費者分類 × 製品分類の度数表における独立性検証 最短距離法 (続き)

非類似度			製品分類									
			χ^2 順位					$\chi^2/d.f.$				
			10 分類	15 分類	20 分類	25 分類	30 分類	10 分類	15 分類	20 分類	25 分類	30 分類
消費者分類	M9 Yules Q	10 分類	11	11	12	12	12	1	1	1	1	1
		15 分類	11	11	12	12	12	1	1	1	1	1
		20 分類	12	11	12	11	11	1	1	1	1	1
		25 分類	12	11	11	12	11	1	1	1	1	1
		30 分類	11	11	11	11	11	1	1	1	1	1
	M10 Russel and Rao	10 分類	2	2	2	2	2	3	3	4	4	3
		15 分類	2	2	2	2	2	4	4	3	3	3
		20 分類	2	2	2	2	2	3	3	3	3	3
		25 分類	6	2	2	2	2	3	3	3	3	3
		30 分類	2	2	2	2	2	3	3	3	3	3
	M11 Chi	10 分類	1	1	1	1	1	11	11	10	10	9
		15 分類	1	1	1	1	1	9	9	9	8	8
		20 分類	1	1	1	1	1	9	9	8	8	7
		25 分類	1	1	1	1	1	8	7	7	7	7
		30 分類	1	1	1	1	1	8	8	7	7	6
	M12 Rogers and Tanimoto	10 分類	13	13	13	14	13	1	1	1	1	1
		15 分類	12	13	15	15	13	1	1	1	1	1
		20 分類	13	13.5	14	14	14.5	1	1	1	1	1
		25 分類	13.5	13	15	15	15	1	1	1	1	1
		30 分類	14	13	15	14	14	1	1	1	1	1
	M13 Sokal and Sneath 1	10 分類	16	14	15	15	15	1	1	1	1	1
		15 分類	13	14	13	14	15	1	1	1	1	1
		20 分類	14	13.5	13	15	13	1	1	1	1	1
		25 分類	13.5	15	13	14	14	1	1	1	1	1
		30 分類	12	12	14	13	12.5	1	1	1	1	1
	M14 Simple	10 分類	15	15	14	13	14	1	1	1	1	1
		15 分類	14	15	14	13	14	1	1	1	1	1
		20 分類	15	15	15	13	14.5	1	1	1	1	1
		25 分類	15	14	14	13	13	1	1	1	1	1
		30 分類	13	14	13	12	12.5	1	1	1	1	1
	M15 Hammann	10 分類	14	16	16	16	16	1	0	0	0	0
		15 分類	15	16	16	16	16	0	0	0	0	0
		20 分類	16	16	16	16	16	0	0	0	0	0
		25 分類	16	16	16	16	16	0	0	0	0	0
		30 分類	16	16	16	16	16	0	0	0	0	0
	M16 Volume	10 分類	12	12	11	11	11	1	1	1	1	1
		15 分類	16	12	11	11	11	0	1	1	1	1
		20 分類	11	12	11	12	12	1	1	1	1	1
		25 分類	11	12	12	11	12	1	1	1	1	1
		30 分類	15	15	12	15	15	0	1	1	1	1

表 A-1-5. アーティスト名一覧

1. モーニング娘。

No	名称	度数	No	名称	度数
1	松浦亜弥	258	2	モーニング娘。	187
3	後藤真希	136	4	藤本美貴	120
5	ごまっとう(後藤真希、松浦亜弥、藤本美貴)	95	6	7AIR・SALT5・11WATER	66
7	モーニング娘。さくら組	65	8	モーニング娘。おとめ組	63
9	ソニン	60	10	安倍なつみ	57
11	安倍麻美	53	12	メロン記念日	49
13	ミニモニ。	47	14	カントリー娘。に紺野と藤本(モーニング娘。)	39
15	カントリー娘。に石川梨華	32	16	ミニモニ。と高橋愛+4KIDS	31
17	ROMANS	31	18	ミニハムず	28
19	ZYX	28			

2. J ポップ

No	名称	度数	No	名称	度数
1	CHEMISTRY	760	2	中島美嘉	601
3	B'z	532	4	EXILE	490
5	SMAP	432	6	宇多田ヒカル	413
7	平井堅	412	8	桑田佳祐	407
9	Mr.Children	379	10	サザンオールスターズ	348
11	MINMI	318	12	MISIA	313
13	一青窈	305	14	森山直太朗	302
15	I WISH	272	16	福山雅治	267
17	aiko	255	18	RUI(柴咲コウ)	242
19	元ちとせ	224	20	鬼束ちひろ	211
21	The Gospellers	198	22	矢井田瞳	195
23	BEGIN	181	24	Kinki Kids	178
25	UA	177	26	ボルノグラフィティ	175
27	山崎まさよし	171	28	ゆず	170
29	アニメ	159	30	光永亮太	152
31	BUMP OF CHICKEN	152	32	水室京介	151
33	久保田利伸	143	34	スガシカオ	136
35	TAK MATSUMOTO	135	36	山下達郎	130
37	中島みゆき	124	38	Sowelu	123
39	林明日香	117	40	GLAY	116
41	Skoop On Somebody	107	42	松任谷由実	98
43	TOKIO	97	44	稲葉浩志	89
45	横原敬之	86	46	V6	81
47	竹内まりや	80	48	嵐	78
49	ゲーム・ミュージック	76	50	スピッツ	74
51	片瀬那奈	74	52	コブクロ	69
53	Ruppina	69	54	長渕剛	68
55	RAG FAIR	67	56	CHAGE & ASKA	64
57	TUBE	61	58	cune	61
59	DREAMS COME TRUE	61	60	小柳ゆき	60
61	くず	59	62	桜庭裕一郎(長瀬智也)	59
63	3B LAB.	57	64	浜田省吾	56
65	布袋寅泰	55	66	松たか子	55
67	柴咲コウ	54	68	中森明菜	53
69	ISSA	53	70	サッズ(Sads/SADS)	53
71	BREATH	50	72	柴田淳	50
73	タッキー&翼	49	74	LOVE PSYCHEDELICO	45
75	YMO(イエロー・マジック・オーケストラ)	43	76	鈴木雅之	42
77	小泉今日子	42	78	徳永英明	40
79	KOKIA	39	80	時給800円	38
81	J-FRIENDS	38	82	CHARCOAL FILTER	37
83	キンモクセイ	34	84	YAIKO(矢井田瞳)	34
85	大滝詠一	33	86	川嶋あい	32
87	Lyrical(露崎春女)	30	88	斉藤和義	30
89	DA PUMP	30	90	坂本真綾	29
91	星村麻衣	27	92	Kiroro	27
93	高橋真梨子	25	94	BUCK-TICK	25
95	THE BOOM	25	96	MY LITTLE LOVER	25
97	川口大輔	24	98	安全地帯	24
99	DEEN	24	100	森田童子	23
101	SAYAKA	21	102	山口百恵	20
103	SING LIKE TALKING	20	104	藤井フミヤ	20
105	佐藤竹善	20	106	JUDY AND MARY	20
107	小島麻由美	20			

3. J 女性ボーカル

No	名称	度数	No	名称	度数
1	浜崎あゆみ	760	2	Every Little Thing	522
3	Do As Infinity	321	4	倉木麻衣	286
5	倅田来未	230	6	day after tomorrow	223
7	島谷ひとみ	178	8	愛内里菜	170
9	ZONE	108	10	SPEED	97
11	Tommy february6	90	12	上戸彩	82
13	the brilliant green	57	14	YUKI	56
15	shela	52	16	TAK MATSUMOTO feat. ZARD	50
17	move	49	18	GARNET CROW	47
19	hitomi	45	20	Yellow Generation	40
21	Tommy heavenly6	37	22	Hysteric Blue	37
23	上原あすみ	32	24	TAK MATSUMOTO feat. 倉木麻衣	32
25	大黒摩季	32	26	ZARD	31
27	PUFFY	30	28	hiro	28
29	織田裕二	27	30	dream	26
31	三枝夕夏 IN db	26	32	相川七瀬	25
33	BON - BON BLANCO	23	34	Folder 5	21
35	推定少女	21			

4. J ビジュアル系

No	名称	度数	No	名称	度数
1	globe	200	2	L'Arc~en~Ciel	134
3	HYDE (hyde)	104	4	玉置成実	90
5	Janne Da Arc	90	6	Gackt	84
7	w-inds.	60	8	T.M.Revolution	56
9	SOPHIA	54	10	Vivian Or Kazuma	43
11	See - Saw	40	12	TETSU69	37
13	石井竜也	33	14	Cyber X feat. Tomiko Va n	29
15	Dir en grey	27	16	TM NETWORK	27
17	PIERROT	23	18	TK PRESENTS (TETSUYA KOM URO)	21
19	S.O.A.P	21	20	Cyber X feat. misono	21
21	Cyber X feat. KEIKO	20	22	KEIKO	20

5. J ヒップホップ

No	名称	度数	No	名称	度数
1	ケツメイシ	402	2	KICK THE CAN CREW	335
3	Dragon Ash	334	4	RIP SLYME	325
5	SOUL 'd OUT	269	6	Crystal Kay	246
7	SUITE CHIC	189	8	DOUBLE	161
9	インターナショナル・レゲエ〜	150	10	Soulhead	146
11	m-flo	139	12	ZEEBRA (ジブラ)	130
13	安室奈美恵	129	14	オレンジレンジ	120
15	Dragon Ash、麻波25、SOURCE	114	16	PUSHIM	106
17	AI	74	18	Crystal Kay loves M-FLO	70
19	キングギドラ	69	20	MOOMIN	66
21	LISA (m-flo/元LiSA)	61	22	FIRE BALL	57
23	M-FLO loves Crystal Kay	57	24	RHYMESTER	57
25	Heartsdales	52	26	m-flo loves melody. & 山本領 平	51
27	PALM DRIVE	48	28	麻波25	38
29	MISS MONDAY	37	30	Sugar Soul	35
31	MIGHTY CROWN	33	32	インディア・アリー	31
33	Retro G-Style	30	34	Keyco	27
35	Home Grown	26	36	HI-D	26
37	アルファ	25	38	F.O.H (Full Of Harmony)	24
39	aki-la	24	40	エレファント・マン	22
41	KAZAMI	21	42	日之内絵美	21
43	KICK THE CAN CREW feat. CASSET	21			

6. J ラップ

No	名称	度数	No	名称	度数
1	S - WORD	64	2	OZROSAURUS	49
3	餓鬼レンジャー	39	4	SPHERE of INFLUENCE	39
5	K - DUB SHINE	33	6	童子 - T	30
7	XBS	28	8	DABO	27
9	AQUARIUS (DELI & YAKKO)	23	10	MURO	23
11	DJ MASTERKEY (MASTERKEY)	22	12	DOBERMAN INC	22
13	DJ YUTAKA	21	14	MACKA - CHIN	21
15	UZI (ウヂ)	21			

7. J パンク

No	名称	度数	No	名称	度数
1	ロードオブメジャー	415	2	HY	389
3	175R	371	4	FLOW	211
5	MONGOL 800	178	6	GOING STEADY	132
7	SHAKALABBITTS (SHAKA LABBITTS)	117	8	ガガガSP	96
9	HAWAIIAN 6	93	10	B - DASH	91
11	THE STAND UP	84	12	太陽族	81
13	NICOTINE	58	14	POTSHOT	58
15	10 - FEET	53	16	JAPAHARINET	52
17	STANCE PUNKS	43	18	BEAT CRUSADERS	38
19	BBQ Chickens	36	20	耳切坊主 (みみちりぼうず)	33
21	Hi - STANDARD	32	22	ELLEGARDEN	31
23	向風	31	24	LOW IQ 01	29
25	オナニーマシーン	28	26	BIVATTCHEE	26
27	SHACHI	25	28	MOLMOTT	24
29	HIGHWAY 61	24	30	CHANGE UP	22
31	Mustang !!	22	32	SABOTEN	22
33	the band apart	21	34	HEAVY HITTER & The Friends	21
35	DOMINO 88	21	36	藍坊主	20

8. 脱ジャンル

No	名称	度数	No	名称	度数
1	エミネム	400	2	女子十二楽坊	216
3	ザ・ビートルズ	180	4	氣志團	152
5	スティーヴィー・ワンダー	149	6	エリック・クラプトン	139
7	ボン・ジョヴィ	119	8	シェリル・クロウ	118
9	エイジア・ダブ・ファウンデーション	116	10	HALCALI	115
11	マイケル・ジャクソン	112	12	GO! GO! 7188	105
13	サンタナ	104	14	スティング	101
15	リラクゼーション / ヒーリング	90	16	夏川りみ	85
17	THE HIGH - LOWS	83	18	シュガー・レイ	78
19	オアシス	77	20	オーディオスレイヴ	76
21	LIV	75	22	ビョーク	75
23	他 : クラシック	70	24	アース・ウインド & ファイアー	68
25	Fayray	64	26	SAKURA	63
27	ダモーン	61	28	アメリカン・ハイファイ	59
29	ザ ローリング ストーンズ	57	30	RIZE	55

8. 脱ジャンル (続き)

No	名称	度数	No	名称	度数
31	レッド・ツェッペリン	55	32	ラス・ケチャップ	55
33	小田和正	54	34	アンドリュウ W . K .	54
35	はなわ	53	36	キャロル	52
37	ウルフルズ	52	38	THE BLUE HEARTS	51
39	エルヴィス・コストロ	47	40	ジェット	47
41	パール・ジャム	47	42	グリーン・デイ	47
43	S O F F e t	45	44	バックストリート・ボーイズ	45
45	サラ・ブライトマン	44	46	ジェイソン ムラーズ	43
47	クラッシュ	42	48	トータス松本	42
49	YOSHII LOVINSON	41	50	BACK DROP BOMB	40
51	マルーン 5	39	52	坂本龍一	38
53	ジェフ・ベック	38	54	モーター・エース	38
55	ボブ マーリィ&ザ ウェイラーズ	37	56	ジェーンズ・アディクション	37
57	Kemuri	37	58	s p e e n a	35
59	スティーリー・ダン	34	60	ジョン・レノン	34
61	Sound Schedule	34	62	ヴァネッサ・カルトン	34
63	ロビー ウィリアムス	34	64	ファウンティンズ オブ ウェイン	34
65	井上陽水	33	66	オリビア・ニュートン・ジョン	33
67	ピリケン	33	68	ヤー ヤー ヤーズ	33
69	他：ボンド	33	70	カーペンターズ	32
71	イーグルス	32	72	デビッド・ボウイ	31
73	エルヴィス・プレスリー	31	74	アイアン・メイデン	31
75	ジャスティンケース	31	76	Z Z	31
77	ベット・ショップ・ボーイズ	30	78	ダリル・ホール&ジョン・オーツ	30
79	R . E . M .	29	80	Amy - N - R y o o	29
81	J U N S K Y W A L K E R S	29	82	綾戸智絵	29
83	バングルス	28	84	アバ	28
85	デイズニー	28	86	明星 (A k e b o s h i)	28
87	m e l o d y .	28	88	S E A L	28
89	システム F	27	90	キザニア・ジョーンズ	27
91	シャギー	27	92	ニュー・オーダー	26
93	エルトン・ジョン	26	94	ポール・マッカートニー	26
95	MONDO GROSSO feat . K j	26	96	B E C K	26
97	尾崎豊	25	98	THE BACK HORN	25
99	ジャネット ジャクソン	25	100	ジョン・メイヤー	25
101	ジ・エクシーズ	25	102	ダナ ・グロヴァー	25
103	マッチボックス・トゥエンティ	25	104	スマッシュ・マウス	25
105	フィル・コリンズ	24	106	タヒチ 8 0	24
107	O R I E N T A L S P A C E	24	108	N a o	24
109	器：フジ子・ヘミング (p)	24	110	ハロウィン	23
111	B @ b y S o u l	23	112	ジャック ジョンソン	23
113	ボーイズ I I メン	23	114	ダンディ・ウォーホルズ	23
115	ゆらゆら帝国	23	116	チルドレン・オブ・ボドム	23
117	マイルス・デイビス	22	118	マーヴィン・ゲイ	22
119	セックス・ピストルズ	22	120	サイモン&ガーファンクル	22
121	エニグマ	22	122	デスティニーズ・チャイルド	22
123	つじあやの	22	124	シックスペンス・ノン・ザ・リッチャー	22
125	シンディ・ローパー	21	126	エアロスミス	21
127	S R - 7 1	21	128	シュガーベイブス	21
129	F R E E T E M P O	21	130	インナー・サークル	21
131	ケイコ・リー	21	132	イギー・ポップ	20
133	X Z I B I T	20	134	ロットングラフティー	20
135	J I N D O U	20	136	平川地一丁目	20
137	エンヤ	20	138	ジ・オーブ	20
139	ウィーザー	20	140	G T S (グルーヴ ザット ソウル)	20
141	ウィル・スミス	20			

9. 個性派

No	名称	度数	No	名称	度数
1	椎名林檎	333	2	アンダーワールド	224
3	MONDO GROSSO	151	4	東京スカパラダイスオーケストラ	149
5	ASIAN KUNG-FU GENERATION	146	6	ケミカル・ブラザーズ	143
7	THEE MICHELLE GUN ELEPHANT	143	8	クレイジーケンバンド	102
9	ACIDMAN	96	10	マッシュヴ・アタック	90
11	ユニコーン	83	12	orange pekoe	82
13	スーパーカー	72	14	レミオロメン	66
15	wyolica	65	16	BONNIE PINK	62
17	Fantastic Plastic Machine	60	18	くるり	60
19	ジェイムス・ブラウン	55	20	PE'Z	55
21	吉田美和(ドリームズ・カム・トゥルー)	51	22	奥田民生	47
23	GOING UNDER GROUND	39	24	Chara	37
25	O.P.King	34	26	ブライアン・セッツァー・オーケストラ	34
27	bird	33	28	エイフェックス ツイン	32
29	NUMBER GIRL	30	30	ジャンキー XL	29
31	U.N.K.L.E.	29	32	LOSALIOS	28
33	スネオヘアー	28	34	ベースメント ジャックス	28
35	THE BLUE HERB	27	36	JUDE(ユダ)	26
37	カルロス・ジーン	26	38	AIR	26
39	曽我部恵一	24	40	ティム・デラックス	24
41	ダフト パンク	24	42	GRAPEVINE	24
43	YO-KING	24	44	ORIGINAL LOVE	23
45	コーネリアス	23	46	キリンジ	23
47	SKETCH SHOW	22	48	石野卓球	22
49	ブライアン・セッツァー	21	50	琉球アンダーグラウンド	21
51	Polaris	20	52	エレファントカシマシ	20
53	DJ シャドウ	20	54	ファットボーイ・スリム	20

10. 洋 正統派ロック

No	名称	度数	No	名称	度数
1	レッド・ホット・チリ・ペッパーズ	367	2	ニルヴァーナ	318
3	レディオヘッド	208	4	U2	138
5	ブラー	126	6	フー・ファイターズ	108
7	ザ・ストロークス	99	8	ズワン	88
9	マンドゥ・ディアオ	88	10	ブライマル・スクリーム	81
11	ザ・ホワイト・ストライプス	67	12	ステレオフォニックス	60
13	ザ・リパティーンズ	59	14	コールドプレイ	56
15	ミューズ	55	16	トラヴィス	55
17	ザ・ダットサンズ	51	18	クーラ・シェイカー	49
19	ザ・ミュージック	48	20	ザ・ジーヴァズ	37
21	MEW	35	22	ザ・コーラル	33
23	スマッシング・パンプキンズ	26	24	ザ・ヴァインズ	22
25	ジェット	22	26	ザ・クーバー・テンブル・クロース	20

11. 洋 女性ボーカル

No	名称	度数	No	名称	度数
1	t . A . T . u .	634	2	アヴリル・ラヴィーン	470
3	エヴァネッセンス	320	4	ミシェル・ブランチ	264
5	マドンナ	258	6	タリア	220
7	ブルー	193	8	ホリー・ヴァランス	191
9	ブリトニー・スピアーズ	187	10	クリスティーナ・アギレラ	153
11	シャナイア・トゥエイン	125	12	ヒラリー・ダフ	114
13	P ! N K	103	14	リリックス	99
15	L U C Y W O O D W A R D	77	16	フェイス・ヒル	75
17	ジャスティン・ティンバーレイク	72	18	アリシア・キーズ	71
19	カイリー・ミノーグ	70	20	セリーヌ・ディオーン	65
21	ケリー・クラークソン	55	22	ダイド	48
23	エイス・オブ・ベイス	47	24	ジュエル	46
25	カーディガンズ	45	26	ウエストライフ	37
27	ノー・ダウト	34	28	スウィートボックス	34
29	ニック・カーター	23	30	ジェシカ・シンブソン	22
31	ナチュラル	21	32	リッキー・マーティン	20

12. 洋 ブラック/ソウル

No	名称	度数	No	名称	度数
1	T L C	428	2	ビヨンセ	405
3	アシャンティ	319	4	マライア・キャリー	192
5	メアリー J . ブライジ	183	6	R . ケリー	171
7	クレイグ・デイヴィッド	155	8	メイシー・グレイ	135
9	フーージェズ	123	10	ホイットニー・ヒューストン	116
11	アリーヤ	112	12	デズリー	105
13	ミスティーク	97	14	ケリー・ローランド	94
15	ニヴェア	93	16	マイア	86
17	エリカ・パドゥ	78	18	トニ・ブラクストン	69
19	ノディーシャ	65	20	スピーチ	56
21	ブルー・カントゥレル	51	22	モニカ	47
23	K - C i & J o J o	46	24	ブラックストリート	43
25	ウェイン ワンダー	41	26	アイシス	40
27	エイメリー	40	28	ソランジュ	37
29	3 L W	36	30	ブライアン・マックナイト	36
31	ロミオ	34	32	ダ プラット	34
33	デボラ・コックス	33	34	ケリー・ブライス	31
35	ジョー	30	36	タイムズ4	23
37	ヘザー・ヘッドリー	22	38	プリンス	21
39	リザ・マフィア	21			

13. 洋 ヒップホップ

No	名称	度数	No	名称	度数
1	ミッシェル・エリオット	193	2	ネリー	182
3	ショーン・ボール	175	4	5 0 セント	169
5	J A Y - Z	144	6	ジャ・ルール	140
7	2バック	117	8	スヌープ・ドッグ(スヌープ・ドギー・ドッグ)	117
9	リル・キム	108	10	N A S	88
11	ブラック・アイド・ピーズ	72	12	D M X	68
13	オービー トライス	60	14	ザ・ルーツ	49
15	ギャング スター	46	16	チンギー	45
17	L . L . クール J	44	18	The Neptunes	44
19	アウトキャスト	44	20	コモン(コモンセンス)	44
21	M s . ジェイド	41	22	バスタ・ライムス	40
23	D r . D r e	37	24	ファット・ジョー	37
25	イヴ	30	26	リュダクリス	30
27	R U N D . M . C .	27	28	ファボラス	24
29	G - ユニット	22	30	M . O . P .	21
31	ジュラシック 5	20	32	フリーウェイ	20
33	エレクトリック	20			

14. 洋 ハードロック

No	名称	度数	No	名称	度数
1	リンキン・パーク	435	2	リンプ・ビズキット (リンプ・ビズキット)	204
3	マリリン・マンソン	143	4	レイジ・アゲインスト・ザ・マシーン	82
5	ニッケルバック	79	6	KORN	59
7	ステインド	57	8	システム・オブ・ア・ダウン	48
9	クレイジー・タウン	43	10	フー・バスターズ	42
11	パドル・オブ・マッド	37	12	デフトーンズ	37
13	ナンバー・ワン・サン	36	14	サード・アイ・ブラインド	33
15	インソレンス	30	16	マーズ・ヴォルタ	30
17	(h e d) P L A N E T E A R T H	25	18	スリップノット	23
19	タッブルート	22			

15. 洋 パンク

No	名称	度数	No	名称	度数
1	S U M 4 1	284	2	グッド・シャーロット	192
3	オフスプリング	181	4	ゼブラヘッド	152
5	ランシド	109	6	アタリス	108
7	メスト	108	8	B L I N K 1 8 2	96
9	シュガー・カルト	91	10	N O F X	81
11	ドゥ・ノッツ	79	12	バステッド	77
13	A L L i S T E R	76	14	G O B	72
15	ニュー・ファウンド・グローリー	55	16	オール・アメリカン・リジェクツ	54
17	ザ・リヴィング・エンド	52	18	シンブル・ブラン	45
19	M x P x	42	20	ザ・ボムジャックス	38
21	ベニー・ワイズ	38	22	ドロップキック・マーフィーズ	35
23	オーソリティ・ゼロ	32	24	ラグワゴン	32
25	ザ・ドナス	31	26	T H E U S E D	30
27	Y e l l o w c a r d	29	28	サムシング・コーボレイト	26
29	ホット・アクション・コップ	25	30	レス・ザン・ジェイク	25
31	スナッフ	23	32	サブライム	23
33	ミー・ファースト&ザ・ギミ・ギミズ	22	34	U S E L E S S I D	21